



Essays on Asset Pricing and Econometrics

Citation

Jin, Tao. 2014. Essays on Asset Pricing and Econometrics. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12269843>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Essays on Asset Pricing and Econometrics

A dissertation presented

by

Tao Jin

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2014

© 2014 Tao Jin

All rights reserved.

Essays on Asset Pricing and Econometrics**Abstract**

This dissertation presents three essays on asset pricing and econometrics. The first chapter identifies rare events and long-run risks simultaneously from a rich data set (the Barro-Ursúa macroeconomic data set) and evaluates their contributions to asset pricing in a unified framework. The proposed model of rare events and long-run risks is estimated using a Bayesian Markov-chain Monte-Carlo method, and the estimates for the disaster process are closer to the data than those in the previous studies. Major evaluation results in asset pricing include: (1) for the unleveraged annual equity premium, the predicted values are 4.8%, 4.2%, and 1.0%, respectively; (2) for the Sharpe ratio, the values are 0.72, 0.66, and 0.15, respectively.

The second chapter, coauthored with Robert J. Barro, estimates the coefficient of relative risk aversion, γ , by exploring the influence of rare disasters on the equity premium. The premium depends on the probability and size distribution of disasters, gauged by proportionate declines in per capita consumption or gross domestic product. Long-term national-accounts data for 36 countries provide a large sample of disasters of magnitude 10% or more. A power-law density provides a good fit to the size distribution. The observed premium of 5% generates an estimated γ close to 3, with a 95% confidence interval of 2 to 4. The results are robust to uncertainty about the values of the disaster probability

and the equity premium.

The third chapter studies the estimation and testing of ARMA(1, 1) models with root cancellation using a new method called “global approach.” With this approach, it shows the asymptotic distributions of the maximum likelihood estimator, gives a complete classification of asymptotic identification categories for all the drifting sequences of parameters, and reveals how the strength of identification of parameters change with the sample size and the sum of the autoregressive (AR) and moving average (MA) parameters. A novel statistic is proposed for conducting joint tests on the AR and MA parameters, which is straightforward to calculate and has some desirable properties.

Contents

Abstract	iii
Acknowledgments	x
1 Rare Events and Long-Run Risks for Asset Prices: Empirical Identification and Evaluation	1
1.1 Introduction	1
1.2 Data and Model	6
1.2.1 Data of Annual C and GDP by Country	7
1.2.2 Model of Rare Events and Long-Run Risks	8
1.2.3 Decomposition of Growth Gaps	16
1.3 Identification of Rare Events and Long-Run Risks	17
1.3.1 Estimation Method	17
1.3.2 Empirical Findings on Rare Events	23
1.3.3 Empirical Findings on Long-Run Risks	33
1.3.4 Distinctions between Rare Events and Long-Run Risks	36
1.3.5 Comparison of Models	41
1.4 Evaluation in Asset Pricing	44
1.4.1 Values of IES ψ , CRRA γ , and DF β	44
1.4.2 Empirical Evaluation	48
1.5 Conclusion	55
2 On the Size Distribution of Macroeconomic Disasters	57
2.1 The Equity Premium in a Model with Rare Disasters	59
2.2 Single-Power-Law Distribution	65
2.3 Double-Power-Law Distribution	72
2.4 Variations in Disaster Probability, Target Equity Premium, and Threshold . .	78
2.5 Can the Equity Premium Be Infinite?	83
2.6 Summary of Main Findings	87
3 Estimation and Testing of ARMA Models with Root Cancellation	88
3.1 Introduction	88

3.2	MLE in the ARMA(1,1) Model	97
3.2.1	Preliminaries	97
3.2.2	Quasi-Likelihood Function	102
3.2.3	A- and B- Series and Uniform Convergence/Boundedness in Probability	105
3.2.4	First Order Conditions	110
3.2.5	Gaussian Processes and Convergence in Finite-Dimensional Distribu- tions	117
3.2.6	Asymptotic Distributions	119
3.3	Hypothesis Testing	123
3.3.1	J-Test	123
3.3.2	Simulation Results	124
3.4	Conclusion	135
References		138
Appendix A Appendix to Chapter 1		144
A.1	Missing Data when $t = t_{i,0}$	144
A.2	Specification of Four Simulation Chains	145
A.3	An Auxiliary Treatment	145
A.4	Conversion to Monthly AR(1) Model	147
A.5	Model Extension: Stochastic Volatility	149
A.6	Supplemental Figures	150
Appendix B Appendix to Chapter 3		157
B.1	Propositions	157
B.2	Proofs	160

List of Tables

1.1	Prior Distributions of Parameters	22
1.2	Facts and Estimates about Rare-Event/Disaster Process	25
1.3	Rare-Event and Long-Run-Risk Parameters in the RE+LRR Model	28
1.4	Comparison to BY's and BKY's Parameter Values	35
1.5	The Mean, Standard Deviation, and First-order Autocorrelation of the RE, LRR, and Noise Components, and the Demeaned Growth Gaps for C and GDP of the US and the World	39
1.6	Correlation Coefficients between the RE, LRR, and Noise Components, and the Demeaned Growth Gaps for C and GDP of the US and the World	40
1.7	Rare-Event Parameters in the RE Model	43
1.8	Long-Run-Risk Parameters in the LRR Model	44
1.9	Asset Pricing Statistics: Data	47
1.10	Asset Pricing Statistics: Data and Model Values for Various CRRA γ	49
1.11	Asset Pricing Statistics: Data and Model Values for Various IES ψ	53
1.12	Asset Pricing Statistics: Data and Predicts from the Three Models	54
2.1	Single and Double Power Laws; Threshold is $z_0 = 1.105$	69
2.2	Single and Double Power Laws with Higher Threshold, $z_0 = 1.170$	82
3.1	Asymptotic Identification Categories for θ_T in ARMA(1,1) Models	93

List of Figures

1.1	$\sigma_{z,k,*}$ and $\sigma_{x,k,*}$ as Functions of k	20
1.2	Probability of World Events	26
1.3	Ex Ante Rare-Event Distribution (C)	30
1.4	Ex Ante Rare-Event Distribution (GDP)	30
1.5	Ex Ante Disaster Distribution (C)	31
1.6	Ex Ante Disaster Distribution (GDP)	31
1.7	Response of $\log C$ to a Typical Rare Event	32
1.8	Estimates for the United States (C)	37
1.9	Estimates for the United States (GDP)	37
1.10	Demeaned Growth Gap Decompostioin for United States (C)	38
1.11	Demeaned Growth Gap Decompostioin for United States (GDP)	38
2.1	Histogram of the empirical density of transformed C disasters and the density estimated by the single power law	63
2.2	Histogram of the empirical density of transformed GDP disasters and the density estimated by the single power law	64
2.3	Estimated log-scale tail distribution and log of transformed ranks of C disaster sizes versus log of transformed C disaster sizes	70
2.4	Estimated log-scale tail distribution and log of transformed ranks of GDP disaster sizes versus log of transformed GDP disaster sizes	71
2.5	Histogram of the empirical density of transformed C disasters and the density estimated by the double power law ($z_0 = 1.105$)	75
2.6	Estimated log-scale tail distribution and log of transformed ranks of C disaster sizes versus log of transformed C disaster sizes	76
2.7	Estimated log-scale tail distribution and log of transformed ranks of GDP disaster sizes versus log of transformed GDP disaster sizes	77
2.8	Estimates of the coefficient of relative risk aversion, γ , for alternative disaster probabilities, given target equity premium of 0.05, threshold of $z_0 = 1.105$, and $\sigma = 0.020$	79

3.1	Boundary Effects on $\hat{\theta}_T$	122
3.2	Coverage Probability of J - and QLR test, $\theta^* = 0$	125
3.3	Coverage Probability of J - and QLR test, $\theta^* = 0.1$	126
3.4	Coverage Probability of J - and QLR test, $\theta^* = -0.2$	126
3.5	Coverage Probability of J - and QLR test, $d^* = 0$	127
3.6	Coverage Probability of J - and QLR test, $d^* = 1$	127
3.7	Coverage Probability of J - and QLR test, $d^* = -3$	128
3.8	Coverage Probability of J - and QLR test, $d^* = 6$	128
3.9	Power of J - and QLR test, $d^* = \theta^* = \theta_0 = 0$	129
3.10	Power of J - and QLR test, $d^* = 1, \theta^* = \theta_0 = 0$	130
3.11	Power of J - and QLR test, $d^* = -2, \theta^* = \theta_0 = 0$	130
3.12	Power of J - and QLR test, $d^* = 4, \theta^* = \theta_0 = 0$	131
3.13	Power of J - and QLR test, $d^* = 4, \theta^* = \theta_0 = -0.2$	131
3.14	Power of J - and QLR test, $d^* = d_0 = 0, \theta^* = 0.1$	132
3.15	Power of J - and QLR test, $d^* = d_0 = 1, \theta^* = 0.2$	133
3.16	Power of J - and QLR test, $d^* = d_0 = -2, \theta^* = 0.4$	133
3.17	Power of J - and QLR test, $d^* = d_0 = -7, \theta^* = -0.5$	134
3.18	Power of J - and QLR test, $d^* = 6, d_0 = 0, \theta^* = 0.3$	134
3.19	Power of J - and QLR test, $d^* = -3, \theta^* = 0.5, \theta_0 = -0.4$	135
A.1	Fitting the Autocovariance Function	148
B.1	Graphs of $h(\theta \phi)$ for Various ϕ	177

Acknowledgments

My greatest appreciation goes to my dissertation advisors—Robert Barro, John Campbell, Anna Mikusheva, and Neil Shephard. I cannot thank Robert enough for his guidance and advice over the years, and this work would not have been possible without all the help I received from him. I am tremendously grateful to John for his kindness, diligence, and brilliance from which I benefited enormously. Anna has always been highly enthusiastic and encouraging, and given me invaluable comments and advice. Neil’s knowledge, insights, and inspiration have been crucial and indispensable for the improvement of this dissertation. I am lucky to have them as my advisors.

I am grateful to James Stock for research meetings that led to Chapter 3 of this dissertation. I thank Donald Andrews, Ian Dew-Becker, Herman van Dijk, Rustam Ibragimov, David Laibson, Emi Nakamura, and Jón Steinsson, among others, for helpful discussions. I also thank seminar participants in the Department of Economics at Harvard University, Harvard Business School, Department of Economics at MIT, MIT Sloan School of Management, PBC School of Finance at Tsinghua University, School of Economics and Finance at The University of Hong Kong, Department of Economics and Finance at City University of Hong Kong, China Economics & Management Academy at Central University of Finance & Economics, School of Economics at Peking University for comments and suggestions.

For my parents, my wife, Shan, and our baby-to-be

Chapter 1

Rare Events and Long-Run Risks for Asset Prices: Empirical Identification and Evaluation

1.1 Introduction

Rare macroeconomic events and long-run risks are two important types of risks in asset markets. A large body of literature has studied their respective roles in asset pricing separately, but few have identified them simultaneously from the data and evaluated their respective contributions to asset pricing in one model. In this study, I address the problem and propose an empirical model of rare events (REs) and long-run risks (LRRs) that provides a legitimate framework for their identification and evaluation. To fully account for the effects of rare macroeconomic events on asset markets, in this study, I use the term “rare events” to include rare disasters and occasional bonanzas.

Similar to previous research, this study treats rare events and long-run risks as unobserved latent variables. It is difficult to separate them, as they are intertwined together in the data. The distinct features of rare events and long-run risks—which I summarize as “sporadic, drastic, and jumping outbursts” and “persistent, moderate, and smooth fluctuations,” respectively—are the foundation for their identification. In addition to their distinctions, knowledge from economic common sense—about event gaps (i.e., the deviation of consumption and output from their potential levels due to current and past rare events) and potential levels of consumption and output—is a necessity for model estimation. Due to the need of incorporating prior information and relative easiness of implementation, the Bayesian Markov-chain Monte-Carlo (MCMC) method is appropriate for estimating the model. The key treatment for identification, elaborated in Section 1.3.1, is to incorporate the aforementioned knowledge into conditional prior beliefs—a novelty of this study.

To estimate the model, I adopt the long-term annual national-accounts data from the Barro-Ursúa macroeconomic data set (Barro and Ursúa, 2010). Observations on per capita consumer expenditure (henceforth, called C) or real per capita gross domestic product (henceforth, called GDP) for 42 economies over 160 years are utilized here. The asset pricing implications of the proposed model are analyzed following a standard treatment, and the representative agents facing an endowment economy are assumed to have the Epstein-Zin-Weil (EZW) preferences (Epstein and Zin (1989); Weil (1990)).

The aim of this study is threefold. First, it proposes a model that integrates rare events and long-run risks, identifies these two types of risks simultaneously from the data, and reveals their distinctions. Empirical estimates disentangle rare events and long-run risks, and reveals three fundamental distinctions—in persistence, volatility, and duration—between

them via the decomposition of growth gaps which are the differences between the actual and the long-term average growth rates.

Second, the proposed model improves upon previous rare-event models' estimation of disaster process. The probability, size, and duration of rare disasters given by the simulation results are close to the facts observed in the data, which indicates that my model well captures those essential elements of the disaster process.

Third, the contributions of the two types of risks in asset pricing are evaluated in a unified framework. If we shut down the rare-event (long-run-risk) channel of the model by restricting the values of the corresponding parameters, we obtain a long-run-risk (rare-event) model. Therefore, the proposed model contains rare-event and long-run-risk models as special cases. Henceforth, I call the three models the "RE+LRR model," "RE model," and "LRR model," respectively. For evaluation in asset pricing, relevant statistics are explored here, including the mean and volatility of the risk-free rate, market return, and equity premium, the Sharpe ratio, and the mean, volatility, first- and second-order autocorrelation of the log price-dividend ratio (PDR) on equity. Using parameter values that match the risk-free rate and the market return observed from the long-term national-accounts data, I calculate the aforementioned asset pricing statistics for the RE+LRR, RE, and LRR models. Major evaluation results include: (1) for the unleveraged annual equity premium, the predicted values are 4.8%, 4.2%, and 1.0%, respectively; (2) for the Sharpe ratio, the values are 0.72, 0.66, and 0.15, respectively.

The rest of the paper is organized as follows. Section 1.2 discusses the data employed in this study and presents the assumptions, settings, and technical details of the model. Section 1.3 explains the estimation method and presents empirical results about rare events,

long-run risks, the distinctions between them, and the comparison of models. Section 1.4 studies the asset pricing implications of the estimated model and evaluates the contributions of rare events and long-run risks to asset pricing statistics. The last section summarizes major findings of this study.

Relation to the literature. The idea of rare events is first introduced to explain the “equity premium puzzle” (Mehra and Prescott, 1985) that the standard consumption-based model with a reasonable coefficient of relative risk aversion (CRRA) generates an equity premium much lower than what people observe from the data. Rietz (1988) and Barro (2006) attribute much of the long-term average equity premium to the risk of rare but disastrous events. In the baseline rare-disaster models, macroeconomic disasters are defined as cumulative decline in C and GDP by about 10% or more (Barro and Ursúa, 2008; Barro and Jin, 2011a). Here I adopt the notion of rare events instead of rare disasters; the latter is used only to examine the accuracy of the empirical estimates. Recently, researchers began to scrutinize the risk of occasional bonanzas, or “rare booms,” and use this notion to explain some stock market behavior (see, e.g., Tsai and Wachter, 2012).

Barro and Ursúa (2008) estimate CRRA γ to be around 3.5 to match the observed equity premium (around 7% on levered equity), using the macroeconomic data of 36 countries, where the size distribution of disasters is gauged by the observed histogram. Barro and Jin (2011a) estimate γ to be around 3 with a 95% confidence interval of 2 to 4, using almost the same data as Barro and Ursúa (2008), with the size distribution of macroeconomic disasters fitted with power laws. As a follow-up to the baseline rare-disaster models that make simplified assumptions about the output and disaster process to obtain a closed-form solution, Nakamura, Steinsson, Barro and Ursúa (2013, NSBU) develop a more realistic empirical

model—the “NSBU model”—to characterize the process of macroeconomic disasters.

The NSBU model improves the baseline rare-event models as follows: (1) allowing for recoveries that usually happen after disasters; (2) letting disasters unfold over multiple years; (3) allowing for correlation in timing of disasters across countries. This model shows that on average a disaster reaches its trough after six years, with a peak-to-trough drop of about 30% in C , and that half of the decline is reversed in a subsequent recovery. Setting intertemporal elasticity of substitution (IES) equal to 2, NSBU find a CRRA of 6.4 to match the average unleveraged equity premium of 4.8% observed from the long-term data of 17 countries. Due to the NSBU model’s improvements upon the baseline rare-event models, the rare-event part of the proposed model in this study adopts NSBU as a benchmark (see Section 1.2.2).

The notion of rare events is also employed by researchers to explain a variety of puzzles and phenomena in asset and foreign exchange markets (e.g. Farhi and Gabaix, 2008; Farhi *et al.*, 2009; Gourio, 2008, 2012; Gabaix, 2012; Wachter, 2013; Seo and Wachter, 2013; Colacito and Croce, 2011a). Barro and Ursúa (2012) provide a detailed review of this literature.

The idea of long-run risks is first introduced in Bansal and Yaron (2004, BY), where the authors maintain that the “risks for the long run,” i.e., small but persistent shocks to expected growth rates and uncertainty, are crucial for explaining various asset market phenomena including the high equity premium, high volatility of stock returns, low value and volatility of the risk-free rate, and the predictability of stock returns. The main results in BY and Bansal, Kiku and Yaron (2012, BKY) are based on a risk aversion of 10 and an IES of 1.5. To characterize the long-run risks, in this study I introduce a slow-varying component of growth rates of consumption and output, which is elaborated in Section 1.2.2

in detail.

Long-run risks are also utilized to explain many other phenomena in asset and foreign exchange markets in subsequent studies, such as Bansal and Shaliastovich (2012), Bansal, Dittmar and Lundblad (2005), Hansen, Heaton and Li (2008), Malloy, Moskowitz and Vissing-Jorgensen (2009), Chen (2010), Colacito and Croce (2011b), Nakamura, Sergeyev and Steinsson (2012). An empirical evaluation of the long-run-risk model is provided by BKY. Meanwhile, Beeler and Campbell (2012) also offer an empirical assessment of the long-run-risk model and document several empirical difficulties for it.

There is a vast literature investigating rare events and long-run risks separately, but, as I have mentioned, few have identified them simultaneously from the data and evaluated their contributions to asset pricing in a unified framework. This study aims to fill the gap of the literature, and proposes an empirical model for the identification and evaluation of the two types of risks.

1.2 Data and Model

In this section, I first introduce the annual C and GDP data used in this study in Section 1.2.1. I then discuss the detailed model setting in Section 1.2.2. Finally, in Section 1.2.3, I derive the decomposition of growth gaps from the constructed model.

For the rare-event part of the model, the main enhancements to the benchmark of NSBU are threefold: (1) To extract more information from the data, I employ a much larger data set with a different data selection criterion. I also model C and GDP in an integrated manner instead of considering C only, or treating C and GDP separately. By integrating the two cases together, I unify the definition of rare events observed in both C and GDP data.

For each country, its long-term average growth rates of C and GDP are assumed to be the same, which excludes the possibility that the consumption-output ratio converges to 0 or ∞ in the long run. (2) To overcome the challenges of identification, I introduce the conditional prior distributions for event gaps and potential levels of consumption (output) that are essential, intuitive and natural. (3) To obtain accurate estimates, I allow the time-varying volatility, carefully address the missing data issue, and introduce one more parameter to the dynamic equation for event gaps to distinguish the “eventful” and “uneventful” periods.

The derivation of the decomposition of growth gaps from the model is critical for analyzing the distinctions and interrelations between rare events and long-run risks. Fundamental distinctions between these two types of risks are revealed by empirical results on the decomposition of (demeaned) growth gaps (presented in Section 1.3).

1.2.1 Data of Annual C and GDP by Country

This study uses the Barro-Ursúa macroeconomic data set (Barro and Ursúa, 2010) which includes the C and GDP series for a total of 42 economies up to 2009. Many C and GDP series are more than 100 years long, which is crucial and required for the estimation of rare-event models. The availability of uninterrupted annual data varies across economies. So, to best utilize the rich information contained in the data set, I adopt the longest possible uninterrupted series between 1850 and 2009 for each economy, yielding a total of 4,696 (5,606) observations for C (GDP). Note that the data set contains more observations. For example, the C series of Sweden dates back to 1800 and the GDP series of the United States dates back to 1790. I choose 1850 as the starting date because it is the earliest year when uninterrupted data are available for at least 10 countries. As the model incorporates the

correlation in the timing of rare events across countries through a world event indicator, it is undesirable if this indicator is estimated by using data from only a few countries. Twelve countries have uninterrupted data available from 1850, including Australia, Belgium, Brazil, Denmark, France, Greece, Netherlands, Norway, Spain, Sweden, the United Kingdom, and the United States.

The difference in data selection criterion and data expansion has made this study's data set much larger than those used in previous studies. For example, in addition to the 17 Organisation for Economic Co-operation and Development (OECD) economies (Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, the United States) and 7 non-OECD economies (Argentina, Brazil, Chile, Mexico, Peru, South Korea, Taiwan) explored in NSBU, 18 more economies are added to this study's data set: Austria, China, Colombia, Egypt, Greece, Iceland, India, Indonesia, Malaysia, New Zealand, Philippines, Russia, South Africa, Singapore, Sri Lanka, Turkey, Uruguay, and Venezuela.

1.2.2 Model of Rare Events and Long-Run Risks

The RE+LRR model contains four main equations: Equation (1.1) for decomposition of consumption and output, Equation (1.4) for growth rates of potential consumption and output, Equation (1.5) for the dynamics of event gaps, and Equation (1.6) for long-run growth rates.¹ Since we are concerned with two cases—C and GDP—an asterisk (*) will denote the general case. For example, when $*$ = C (GDP), $t_{i,0,*}$ means the earliest date when uninterrupted C (GDP) data are available for country i . In general, the panel data

¹Two cases (one for C and one for GDP) exist for each of the four main equations. Equations (1.1), (1.4), and (1.5) are similar to those in the NSBU benchmark model.

for country i are unbalanced, i.e., we should not expect that $t_{i,0,C} = t_{i,0,GDP}$. I define $t_{i,0} = \min\{t_{i,0,C}, t_{i,0,GDP}\}$ to denote the earliest date when uninterrupted C or GDP data are available for country i .

To pool information of rare events and long-run risks, it is assumed that the RE and LRR parameters are constant over time and across countries, while other parameters are allowed to vary across countries and over time.² It is also assumed that the prior distributions of parameters and unobserved quantities are independent.

Decomposition of consumption (output). The observed data on log C (GDP), denoted by $y_{i,t,*}$, is decomposed into three unobserved variables

$$y_{i,t,*} = x_{i,t,*} + z_{i,t,*} + \epsilon_{i,t,*}, \quad (1.1)$$

where $x_{i,t,*}$ is the potential level, or the trend, of log C (GDP), $z_{i,t,*}$ is the “event gap” in C (GDP) characterizing the deviation of log C (GDP) from the potential level due to current and past events, and $\epsilon_{i,t,*}$ is the temporary shock that is independent and identically distributed (i.i.d.) normal.

Unlike NSBU, where the volatility of $\epsilon_{i,t,C}$ (i.e., $\sigma_{\epsilon,i,t,C}$) can take only two values for each country i (one before 1946 and one after), in this study I allow $\sigma_{\epsilon,i,t,*}$ to vary not only across countries but also over time, and it is estimated using *rolling standard deviation* (i.e., *realized volatility*). Here, the window size is taken to be 20 periods. Specifically, in the MCMC estimation, for a given country i , the random draw of $\sigma_{\epsilon,i,t,*}$ is obtained by using the most recent values of $(\epsilon_{i,t-19,*}, \dots, \epsilon_{i,t,*})$ as input. When information about previous

²Rare-event parameters include p_W , p_{bW} , p_{bI} , p_e , $\rho_{z,0,*}$, $\rho_{z,1,*}$, θ_* , $\sigma_{\theta,*}$, ϕ_* , and $\sigma_{\phi,*}^2$; the long-run-risk parameters include λ_* and w_* ; other parameters include β_i , $\sigma_{\epsilon,i,t,*}$, $\sigma_{\eta,i,*}$ and $\sigma_{v,i,*}$. These parameters will be discussed later.

volatilities is insufficient, i.e., when $t_{i,0,*} \leq t < t_{i,0,*} + 19$, the value of $\sigma_{\epsilon,i,t,*}$ is drawn using $(\epsilon_{i,t_{i,0,*},*}, \dots, \epsilon_{i,t_{i,0,*}+19,*})$ as input.

World and country event connection. Similar to NSBU, two types of rare events, i.e., the world events and country events, are connected through two types of binary latent variables: the world rare-event indicator $I_{W,t}$ and the rare-event indicator $I_{i,t}$ for country i . If $I_{W,t}$ takes value 0, the world is in “normal times” at time t ; if it takes value 1, the world is in a rare event. The meaning of $I_{i,t}$ is similar. We assume that

$$\Pr(I_{W,t} = 1) = p_W \quad (1.2)$$

for each period t . The conditional probability $\Pr(I_{i,t} = 1 | I_{i,t-1}, I_{W,t})$ is assumed to be

$$\Pr(I_{i,t} = 1 | I_{i,t-1}, I_{W,t}) = \begin{cases} p_{bI}, & \text{if } I_{i,t-1} = 0 \text{ and } I_{W,t} = 0 \\ p_{bW}, & \text{if } I_{i,t-1} = 0 \text{ and } I_{W,t} = 1 \\ 1 - p_e, & \text{if } I_{i,t-1} = 1 \end{cases} \quad (1.3)$$

for $t > t_{i,0}$. When $t = t_{i,0}$, Formula (1.3) is not directly applicable, as $I_{i,t_{i,0}-1}$ is missing (see Appendix A.1 for the technical treatment of this issue). The “missing data” treatment for the case of $t = t_{i,0}$ in Formulas (A.1) and (A.2) is not present in NSBU. According to Formula (1.3), p_{bW} (p_{bI}) denotes the conditional probability that a rare event begins in a country, given that a (no) world event occurs at the same time and there is no rare event in that country in the previous period; p_e denotes the conditional probability that a country exits the event state, given that a rare event is present in that country in the previous period.

Growth rates of potential consumption (output). Both rare events and long-run risks influence the potential log C (GDP) $x_{i,t,*}$, which evolves according to the following dynamic

equation

$$\Delta x_{i,t,*} = \mu_{i,t,*} + \eta_{i,t,*} + I_{i,t}\theta_{i,t,*}, \quad (1.4)$$

where $\Delta x_{i,t,*} \equiv x_{i,t,*} - x_{i,t-1,*}$ and $\mu_{i,t,*}$ is the slow-varying component of growth rate of C (GDP), $\eta_{i,t,*}$ is the i.i.d. normal permanent shock to potential C (GDP) with a country specific variance $\sigma_{\eta,i,*}^2$ and $\theta_{i,t,*}$ is the permanent shift in the level of potential C (GDP) due to a rare event. The slow-varying component $\mu_{i,t,*}$ characterizes the long-run risks and will be fully discussed later.

Dynamics of event gaps. The event gap $z_{i,t,*}$ is captured by a modified (Autoregressive Distributed Lag) ADL(1,0) model

$$z_{i,t,*} = \rho_{z,I_{i,t},*} z_{i,t-1,*} - I_{i,t}\theta_{i,t,*} + I_{i,t}\phi_{i,t,*} + v_{i,t,*} \quad (1.5)$$

where $\rho_{z,I_{i,t},*}$ is the first-order autoregressive coefficient, $\phi_{i,t,*}$ is the short-run event shock on C (GDP), and $v_{i,t,*}$ is the i.i.d. normal shock to event gap $z_{i,t,*}$ with a country specific variance $\sigma_{v,i,*}^2$.

The permanent effects of rare events on C or GDP (i.e., $\theta_{i,t,*}$) is allowed to be either positive or negative, while the short-run shocks of these events (i.e., $\phi_{i,t,*}$) are assumed to be negative. Unlike the NSBU model which uses one parameter ρ_z , two autoregressive coefficients, $\rho_{z,0,*} \in [0, 1)$ (for $I_{i,t} = 0$) and $\rho_{z,1,*} \in [0, +\infty)$ (for $I_{i,t} = 1$), are introduced to Equation (1.5) to improve the estimation accuracy and give richer information on the process of event gaps.

This treatment is plausible, since the event gap may have different extent of persistency in eventful and uneventful periods. Based on economic intuition, the process of event gaps should be stationary during uneventful periods, so $|\rho_{z,0,*}|$ has to be less than 1. However,

as event gaps may magnify during disasters, no such restriction is imposed for eventful periods in my model.

Long-run growth rates. The slow-varying component $\mu_{i,t,*}$, called the “long-run growth rates,” is introduced to reflect the persistent predictable component of growth rates. Long-run growth rates $\mu_{i,t,*}$ are defined as

$$\mu_{i,t,*} = \beta_i + u_{i,t,*} \quad (1.6)$$

for $t \geq t_{i,0}$, where β_i , the long-term average growth rate of country i , is a country specific constant, and $\{u_{i,t,*}\}_t$ is the process of fluctuations in $\mu_{i,t,*}$. I assume that β_i is the same for both C and GDP cases so as to exclude the situation where the consumption-output ratio approaches to 0 or ∞ in the long run. This is another advantage of integrating C and GDP cases together.

The sequence of shocks $\{u_{i,t,*}\}$ is a Markov process which is of the form

$$u_{i,t,*} = v_*^\Delta \cdot v_{i,t,*}, \quad (1.7)$$

where v_*^Δ is the grid size of shocks and $\{v_{i,t,*}\}_t$ is a Markov chain on \mathbb{Z} , the set of integers.

In this study, I take $v_*^\Delta = 0.01$ for both C and GDP cases. The process $\{v_{i,t,*}\}_t$ is governed by the following transition probabilities

$$\Pr(v_{i,t+1,*} = v_{i,t,*} + k | v_{i,t,*}) = \begin{cases} e^{-\lambda_{i,*}}, & \text{if } k = 0 \\ \frac{\lambda_{i,*}^{|k|} e^{-\lambda_{i,*}}}{|k|!} \frac{w_{i,*}^{|v_{i,t,*}|}}{1 + w_{i,*}^{|v_{i,t,*}|}}, & \text{if } k \neq 0 \text{ and } k(v_{i,t,*}) \leq 0 \\ \frac{\lambda_{i,*}^{|k|} e^{-\lambda_{i,*}}}{|k|!} \frac{1}{1 + w_{i,*}^{|v_{i,t,*}|}}, & \text{if } k \neq 0 \text{ and } k(v_{i,t,*}) > 0 \end{cases}, \quad (1.8)$$

for any $k \in \mathbb{Z}$, where $\lambda_{i,*} > 0$, and $w_{i,*} (> 0)$ is the *weight parameter*. As the transition

probability is adapted from the Poisson distribution with parameter $\lambda_{i,*}$, I call $\lambda_{i,*}$ the *Poisson parameter* in this study.

When $w_{i,*} \in (0, 1)$, $\{v_{i,t,*}\}_t$ is a *mean averting* process, when $w_{i,*} > 1$, a *mean reverting* process, and $w_{i,*} = 1$ a *random walk*. When $w_{i,*} = 1$, the process $\{\mu_{i,t,*}\}_t$ is equivalent to the following unit autoregressive root model (see Stock (1994) for a review on “unit roots, structural breaks and trends”)

$$\mu_{i,t,*} = \beta_i + u_{i,t,*} \quad (1.9)$$

for $t \geq t_{i,0}$ and

$$u_{i,t,*} = u_{i,t-1,*} + v_{i,t,*}^\Delta \quad (1.10)$$

for $t \geq t_{i,0} + 1$, where $v_{i,t,*}$ are i.i.d. and follow a *symmetric two-sided Poisson distribution*

$$\Pr(v_{i,t,*} = k) = \begin{cases} e^{-\lambda_{i,*}}, & \text{if } k = 0 \\ \frac{\lambda_{i,*}^{|k|} e^{-\lambda_{i,*}}}{2 \cdot |k|!}, & \text{if } k \neq 0 \end{cases}. \quad (1.11)$$

Namely, the distribution of $v_{i,t,*}$ is symmetric about zero, and $|v_{i,t,*}|$ follows the Poisson distribution with parameter $\lambda_{i,*}$.

It is counterintuitive for $\{v_{i,t,*}\}_t$ to be a mean averting process. It can be shown that a larger $w_{i,*}$ means either a lower speed of aversion ($w_{i,*} < 1$) or a higher speed of reversion ($w_{i,*} > 1$). When $w_{i,*} > 1$, i.e., $\{v_{i,t,*}\}_t$ is mean reverting, the *reversion level* of $\{v_{i,t,*}\}_t$ is 0, which is also the long-term mean value of $\{v_{i,t,*}\}_t$ no matter where $\{v_{i,t,*}\}_t$ starts.³ Thus, β_i is the long-term mean value of $\{\mu_{i,t,*}\}_t$, i.e., the long-term average growth rate of potential C and GDP for country i with the impacts of rare events excluded. As mentioned

³Rigorously speaking, the “mean reverting” feature is the case only when $v_{i,t,*}$ is deviated from the reversion level 0. When $v_{i,t,*} = 0$, there is some tendency to deviate. However, no matter how far away $v_{i,0,*}$ is from 0, $v_{i,t,*}$ will fluctuate around 0 almost surely when t is sufficiently large.

previously, I assume that $\lambda_{i,*} = \lambda_*$ and $w_{i,*} = w_*$ for each country i . Clearly, a larger λ_* or a smaller w_* corresponds to a higher volatility of $\mu_{i,t,*}$.

Unlike BY where the persistent predictable component x_t is modeled using an (autoregressive) AR(1) process

$$x_{t+1} = \rho x_t + \varphi_e \sigma e_{t+1}, \quad (1.12)$$

this study describes the fluctuating expected growth rates with the Markov process of fluctuations stated above. Yet the two processes are similar overall: both of them characterize small persistent shocks in long-run expected growth naturally and conveniently.

In (1.12), the first-order autoregressive coefficient ρ is very large (close to 0.98) so as to characterize the very large persistence in expected growth rate. This is similar to the mean-reverting case of $w_{i,C} > 1$ with $w_{i,C}$ close to 1 in (1.8). As a comparison, let's see the extreme unit root case in (1.12): With $\rho = 1$, $\{x_t\}_t$ becomes a random walk which resembles the random walk case of $w_{i,C} = 1$ which is discussed earlier. As the discrete symmetric two-sided Poisson distribution is close to a discretization of normal distributions, the Markov process $\{u_{i,t,*}\}_t$ can be viewed approximately as a discretization of $\{x_t\}_t$. Also note that the discrete feature of $\{v_{i,t,*}\}_t$ is not an issue here, as the simulation will be iterated for hundreds of thousand times, and we are interested in quantities like the posterior mean and standard deviation.

I choose the Markov process to model the fluctuating expected growth rates for two reason. First, as a discretized version of the fluctuating expected growth rates, it reflects the intuitive idea that the growth rates are stationary in the short run but can be shifting in the long run. To analyze the asset pricing implications of the proposed model, it is necessary to solve an integral equation on a grid. Therefore, discretizing the model is

inevitable and employing the Markov process is relatively easy. (In Appendix A.5, I list a model where the fluctuating expect growth rates are described by an AR(1) process, similar to the setting in BY.)

Second, even if an AR(1) process were adopted to model the the fluctuating expected growth rates, it would not be feasible to directly compare estimates from an annual model with parameter values in a monthly model, e.g., as in BY and BKY. Thus, the gain from using an AR(1) process can be minimal. In Section 1.3.3, after converting parameters of an annual model to corresponding parameters of the monthly AR(1) model, I compare my estimates of long-run risks with BY's and BKY's parameter values. (For details of this conversion, see Appendix A.4.)

In this study, I allow the volatility of the temporary shocks $\epsilon_{i,t,*}$ (which are estimated using *realized volatility*) to vary over time instead of integrating formally the uncertainty risk (or the volatility risks) in the model—since the stochastic volatility is less significant in annual data (as employed in this study) than in daily or monthly data.⁴ Similar to the autoregressive process, realized volatility is also widely adopted in quantifying time-varying volatility. The introduction of the time-varying volatility makes the model estimation more accurate.

The phenomena of postwar moderation and “great moderation” (Stock and Watson, 2003) further dampen the importance of stochastic volatility in the annual data, as shown by both the data and estimates in this study.⁵ Thus I primarily focus on the fluctuating

⁴Appendix A.4 provides the formula for the autocorrelation function (ACF) of the temporal aggregation of AR(1) series, which gives us a rough idea about how the persistence parameter changes from monthly model to annual model.

⁵Some studies attribute the declines in consumption and output volatility to changes in the measurement of national accounts, e.g., Romer (1986) and Balke and Gordon (1989).

expected growth rates in this study, leaving the uncertainty risk to future research. By comparing the results in this study and future research, we can evaluate the relative importance of the fluctuating expected growth rates and the uncertainty risk in the annual data. In Appendix A.5, I list an empirical model for future research where the long-run-risk part is adapted from BY's setting. The main extension there is to formally integrate the stochastic volatility.

1.2.3 Decomposition of Growth Gaps

As both rare events and long-run risks influence the growth of consumption and GDP, it is key to decompose the growth rates into the corresponding components and examine their performance.

For country i and $t > t_{i,0,*}$, define the *growth gap* $\widetilde{\Delta y_{i,t,*}}$ (of C or GDP) as the difference between the actually growth rates and the long-term average growth rate β_i , i.e.,

$$\widetilde{\Delta y_{i,t,*}} \triangleq \Delta y_{i,t,*} - \beta_i = y_{i,t,*} - y_{i,t-1,*} - \beta_i,$$

then it can be decomposed into three components as follows

$$\widetilde{\Delta y_{i,t,*}} = \text{RE}_{i,t,*} + \text{LRR}_{i,t,*} + \text{N}_{i,t,*},$$

where

$$\text{RE}_{i,t,*} \triangleq I_{i,t,*}\theta_{i,t,*} + \Delta z_{i,t,*} = I_{i,t,*}\theta_{i,t,*} + z_{i,t,*} - z_{i,t-1,*}$$

is the *rare-event component*,

$$\text{LRR}_{i,t,*} \triangleq \mu_{i,t,*} - \beta_i$$

is the *long-run-risk component*, and

$$N_{i,t,*} \triangleq \eta_{i,t,*} + \Delta\epsilon_{i,t,*} = \eta_{i,t,*} + \epsilon_{i,t,*} - \epsilon_{i,t-1,*}$$

is the *noise/random shocks* unexplained by the model. The long-term mean value of $LRR_{i,t,*}$ and $N_{i,t,*}$ is 0, while that of $RE_{i,t,*}$ is not. Let $RE_{i,t,*}^{DM}$ denote the *demeaned* $RE_{i,t,*}$, and

$$\Delta y_{i,t,*}^{DM} \triangleq RE_{i,t,*}^{DM} + LRR_{i,t,*} + N_{i,t,*}$$

denote the *demeaned growth gap*. All the components of the (demeaned) growth gaps will be identified after the proposed model is estimated.

1.3 Identification of Rare Events and Long-Run Risks

In this section, I first discuss the main technical challenges and the model estimation method in Section 1.3.1. Then in Section 1.3.2–1.3.5, I present the empirical findings on rare events, long-run risks, the distinctions between these two risks, and the comparison of the RE+LRR, RE, and LRR models.

The Bayesian estimates of the parameters for rare events and long-run risks, including the mean and standard deviation of their prior and posterior distributions, are reported in Table 1.3. I take the posterior mean of each parameter as the corresponding (point) estimate.

1.3.1 Estimation Method

In this section, I present my solution to the identification of rare events and long-run risks and discuss prior distributions for parameters and the Bayesian MCMC estimation

procedure in detail.

Bayesian Markov-chain Monte-Carlo

Rare events and long-run risks are shocks of different nature, and the statistical distinctions between them enable us to identify them from data simultaneously. However, realizing their identification is a technical challenge. Bayesian MCMC is an appropriate choice for estimating the model in this study, because: (1) necessary information can be incorporated into prior beliefs, and (2) it is relatively easy to implement for as complicated a model as the one proposed here. The crucial treatment for solving the identification problem is the introduction of the conditional prior distributions of $z_{i,t,*}$ and $x_{i,t,*}$. For more accurate estimates, an auxiliary technical treatment is adopted in the estimation procedure (see Appendix A.3).

Conditional prior distribution of event gaps. By economic common sense, it is intuitively clear that event gaps will gradually diminish if no events occur in a country. Based on this fact, I specify the conditional prior distribution of $z_{i,t,*}$ as follows. When $I_{i,t} = 1$, i.e., country i is in a rare event at time t , the prior distribution of $z_{i,t,*}$ is assumed to be $\mathcal{N}(0, \sigma_{z,0,*}^2)$. I take $\sigma_{z,0,*} = 2$ which is very large, so the prior is fairly uninformative on a region local to 0. Suppose year t is the first uneventful year after a rare event in country i , then (1.5) becomes

$$z_{i,t,*} = \rho_{z,0,*} z_{i,t-1,*} + v_{i,t,*},$$

which implies that

$$\begin{aligned}\text{Var}(z_{i,t,*}) &\leq (\rho_{z,0,*} \cdot \text{S.D.}(z_{i,t-1,*}) + \text{S.D.}(v_{i,t,*}))^2 \\ &\leq (0.9 \cdot \sigma_{z,0,*} + \max(\sigma_{v,i,*}))^2,\end{aligned}\tag{1.13}$$

i.e.,

$$\text{S.D.}(z_{i,t,*}) \leq \sigma_{z,1,*} \triangleq 0.9 \cdot \sigma_{z,0,*} + 0.02 = 1.82,$$

where “S.D.” stands for “standard deviation.” When year t is the k -th uneventful year after the most recent rare event in country i , the upper bound $\sigma_{z,k,*}$ of $\text{S.D.}(z_{i,t,*})$ can be calculated recursively, and I assume that the prior distribution of $z_{i,t,*}$ follows $\mathcal{N}(0, \sigma_{z,k,*}^2)$.⁶ Note that the above specification of prior distributions of event gap $z_{i,t,*}$ is natural and intuitively obvious, and is conditional on when the last event before year t happens in country i .

Conditional prior distribution of potential consumption and output. Based on the prior distribution of $z_{i,t,*}$, I derive the conditional prior distribution of $x_{i,t,*}$ as follows. According to Equation 1.1, the upper bound $\sigma_{x,k,*}$ of $\text{S.D.}(x_{i,t,*})$ clearly satisfies

$$\sigma_{x,k,*} \leq \sigma_{z,k,*} + \max(\sigma_{\epsilon,i,t,*}) = \sigma_{z,k,*} + 0.15,\tag{1.14}$$

when year t is the k -th uneventful year after the most recent event in country i . Correspondingly, the prior distribution of $x_{i,t,*}$ is specified to be $\mathcal{N}(y_{i,t,*}, \sigma_{x,k,*}^2)$. Figure 1.1 shows the standard deviation $\sigma_{z,k,*}$ ($\sigma_{x,k,*}$) of the prior distribution of $z_{i,t,*}$ ($x_{i,t,*}$) as a function of k . As k goes to ∞ , $\sigma_{z,k,*}$ ($\sigma_{x,k,*}$) is decreasing and converges to 0.2 (0.35) which is very large based on the economic common sense. So the prior distributions of $z_{i,t,*}$ and $x_{i,t,*}$ are fairly

⁶Here, $k = 0$ indicates that country i is in a rare event. In the simulation, if no event happens in year $t_{i,0}$ for country i , a simple simulation using probability p_i will be implemented to determine the number k .

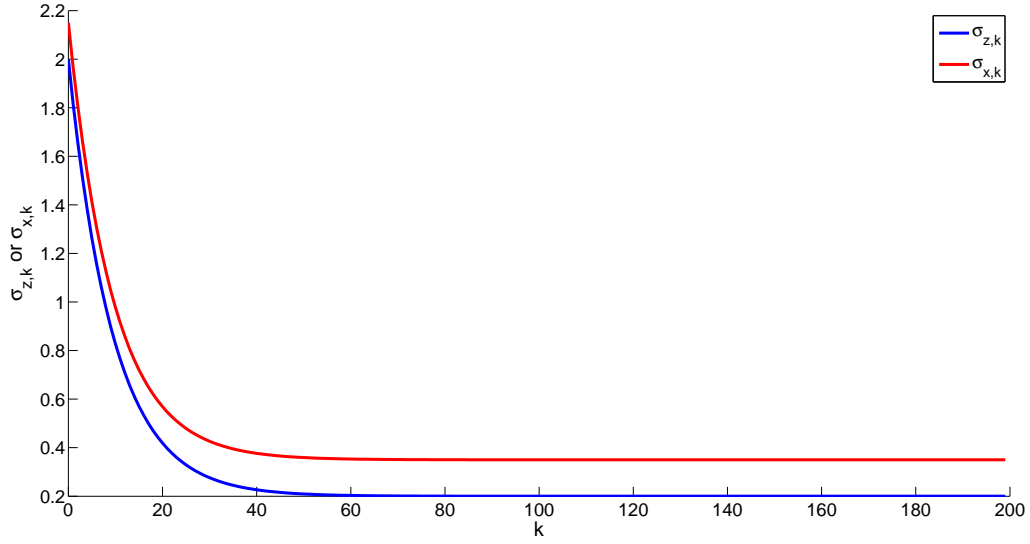


Figure 1.1: $\sigma_{z,k,*}$ and $\sigma_{x,k,*}$ as Functions of k .

uninformative.

In this study, a prior is “uninformative” means that the posterior distribution is proportional to the likelihood. With an uninformative prior, the maximum likelihood estimate should corresponds to the mode of the posterior distribution. Thus, a typical uninformative prior for a parameter is the uniform distribution on an infinite interval (i.e., a half-line or the entire real line). Extending the above idea, I also say the uniform distribution on a finite interval is uninformative, if, according to other information or knowledge, the finite interval contains the parameter with probability 1. More generally, I say a prior distribution is “not very informative,” if it is close to a flat prior. In this study, the general guideline for the specification of priors is to make them as uninformative as possible (on certain regions). Thus, many priors are taken to be uniform. The prior distributions of $z_{i,t,*}$ and $x_{i,t,*}$ are specified as above, and the rest of the prior distributions is elaborated as follows.

Prior Distributions of Parameters

In this study, $\theta_{i,t,*}$ is assumed to follow the normal distribution $\mathcal{N}(\theta_*, \sigma_{\theta,*}^2)$ and $\phi_{i,t,*}$ is assumed to follow the truncated normal distribution $\mathcal{TN}(\phi_*^\diamond, \sigma_{\phi,*}^{\diamond 2}, -\infty, 0)$, where ϕ_*^\diamond and $\sigma_{\phi,*}^{\diamond 2}$ denote the mean and variance, respectively, of the underlying normal distribution (i.e., the normal distribution before truncation). The mean value and standard deviation of $\phi_{i,t,*}$ are denoted by ϕ_* and $\sigma_{\phi,*}$, respectively. Another possible choice for the prior distribution of $\theta_{i,t,*}$ and $\phi_{i,t,*}$ is the exponential distribution. Corresponding to Barro and Jin (2011a), if $z \equiv \frac{1}{1-b} \sim$ power law distribution with (upper-tail) exponent α , where the disaster size b is the fraction of contraction in C (GDP), then $\xi = -\ln z \sim$ exponential distribution with rate parameter α . This relationship suggests exponential distributions for $\theta_{i,t,*}$ and $\phi_{i,t,*}$. For comparing the results here with those in NSBU, I adopt the normal distribution assumption in this study.

The long-term average growth rate β_i of country i is assumed to be

$$\beta_i = b_0 + b^\Delta \cdot b_i,$$

where b_0 is a positive constant, $b^\Delta = 0.01$, and b_i follows the symmetric two-sided Poisson distribution with parameter λ_0 (see Formula (1.11) for the definition of the symmetric two-sided Poisson distribution). Thus, β_i follows a two-sided Poisson distribution with parameters (b_0, λ_0) which is denoted by $\mathcal{TP}(b_0, \lambda_0)$.

Clearly, the mean value of β_i is b_0 and the variance of b_i is $\lambda_0 + \lambda_0^2$. To estimate b_0 and λ_0 , I calculate the long-term average growth rates for all the economies in the data set, and the long-term average growth rate is computed as the mean value of the long-term average growth rate of C and GDP. Using the Barro-Ursúa data set, the mean value of the growth

Table 1.1: Prior Distributions of Parameters

Parameter	Distribution	Parameter	Distribution
$p_W \sim \mathcal{U}(0, 0.1)$		$p_{bI} \sim \mathcal{U}(0, 0.03)$	
$p_{bW} \sim \mathcal{U}(0, 1)$		$p_e \sim \mathcal{U}(0.1, 1)$	
$\theta_* \sim \mathcal{N}(0, 0.4^2)$		$\sigma_{\theta_*} \sim \mathcal{U}(0.01, 0.25)$	
$\phi_*^\diamond \sim \mathcal{U}(-0.25, 0)$		$\sigma_{\phi_*}^\diamond \sim \mathcal{U}(0.01, 0.25)$	
$\rho_{z,0,*} \sim \mathcal{U}(0, 0.9)$		$\rho_{z,1,*} \sim \mathcal{U}(0, 2)$	
$\sigma_{v,i,*} \sim \mathcal{U}(0.001, 0.02)$			
$\lambda_* \sim \mathcal{U}(0.01, 0.5)$		$w_* \sim \mathcal{U}(0.01, 100)$	
$\sigma_{\epsilon,i,t,*} \sim \mathcal{U}(0.001, 0.15)$		$\sigma_{\eta,i,*} \sim \mathcal{U}(0.001, 0.15)$	

rates of 42 economies is 0.0187 and the standard deviation is 0.00551, which corresponds to a variance of 0.304 of b_i , or a λ_0 of 0.244. Therefore, I take $b_0 = 0.02$ and $\lambda_0 = 0.24$ in this study.

Prior distributions for the rest of the parameters are listed in Table 1.1. Except for prior distributions for p_{bI} , $\rho_{z,1,*}$, λ_* and w_* , the prior distributions for other parameters in Table 1.1 are the same as or close to the corresponding specifications in the NSBU model.

Estimation Procedure

The model is estimated by the Bayesian MCMC method, which has been applied to increasingly more problems in economics and finance, e.g., Chib, Nardari and Shephard (2002), Pesaran, Pettenuzzo and Timmermann (2006), Koop and Potter (2007), Stella and Stock

(2013). Specifically, I use the algorithm of the Gibbs sampler for the random draws of parameters and unobserved quantities (see Gelman, Carlin, Stern and Rubin (2004) for a more comprehensive introduction of the MCMC algorithms).

The convergence of MCMC simulation is guaranteed under very general conditions. In order to accurately estimate parameters and unknown quantities, I run four simulation chains, which is similar to the procedure in NSBU (see Appendix A.2 for details of the specification of the four simulation chains). Besides simulating multiple sequences with over-dispersed starting points throughout the parameter space and visually evaluating the trace plots of parameters and unknown quantities from the simulation, I also assess the convergence by comparing variation “between” and “within” simulated sequences (see Gelman, Carlin, Stern and Rubin (2004, Chap. 11) for a discussion of this method).

After 150,000 iterations, the simulation results from the four sets of far-apart initial values stabilize and become very close to each other. So I iterate each chain for 400,000 times and use the later 200,000 iterations to analyze the posterior distributions of parameters and unknown quantities that interest me. The first 200,000 iterations are dropped as burn-in.

1.3.2 Empirical Findings on Rare Events

The empirical findings about rare events fall into two subcategories: (1) The estimates of rare disaster process, including the probability, sizes, and durations of disasters, which are important for a better understanding of disasters. The comparison between the disaster process estimates and the facts observed in data shows that the proposed model matches up with data very well. (2) The estimates of rare-event parameters, which are necessary for the evaluation in asset pricing.

Event/Disaster Process Estimates

The proposed model yields very good estimates for the disaster process overall. Table 1.2 lists facts and estimates about the rare-event/disaster process, including the probability, sizes, and durations of rare events/disasters. I also list the corresponding estimates in NSBU as a reference. Note that NSBU's data set is only a subset of the data set used in this study, so we should not simply compare the three columns in Table 1.2. Despite the differences between the data sets, we still see improvements in disaster process estimates. Now I analyze these findings in detail.

Rare events contain a huge amount of risks. Based on the definition of disasters in Section 1.1 and the use of peak-to-trough measurement, I find 163 episodes of C disasters and 170 GDP disasters in the data.⁷ The corresponding empirical probability for C (GDP) disaster is 3.84% (3.31%), the average peak-to-trough contraction in C (GDP) is 20%, and the average disaster duration is 3.5 years for both C and GDP. Note that there is a 0.5 percentage-point gap in the empirical probability of C and GDP disasters. One reason for this is the difference in the availability of data. There are more observations for GDP than for C, which might affect the characteristics of the data. The gap might have been narrower if there were more observations for both cases, or just for C.

The above facts are close to those documented in Barro and Jin (2011a), where the disaster size threshold is taken to be 9.5%, and 99 C disasters and 157 GDP disasters are found in the long-term national-accounts data for 36 economies. According to Barro and Jin

⁷When determining C (GDP) disaster periods, I allow for temporary recoveries of C (GDP) for one year. That is, two declining periods of $[t_0, t_1]$ and $[t_1, t_2]$ will be merged into one period $[t_0, t_2]$, if the merged period $[t_0, t_2]$ induces a larger peak-to-trough drop than does the period $[t_0, t_1]$. However, if there is a recovery for at least two consecutive years between two declining periods, those two periods won't be merged.

Table 1.2: Facts and Estimates about Rare-Event/Disaster Process

	Model [†]	Data [§]	NSBU [‡]
Event Prob. (World)	0.0586	—	0.037
Event Prob. (Country)	0.0513	—	0.028
Disaster Prob. (C)	0.0331	0.0384	< 0.028
Disaster Prob. (GDP)	0.0336	0.0331	—
Avg. Event Duration	3.2	—	6
Avg. Disaster Duration (C)	3.2	3.5	—
Avg. Disaster Duration (GDP)	3.3	3.5	—
Avg. Drop in a Disaster (C & GDP)	23%	20%	27%
Avg. Event Temp. Drop in a Year (C)	8.2%	—	10.5%
Avg. Event Temp. Drop in a Year (GDP)	8.3%	—	—
Avg. Event Perm. Effect in a Year (C & GDP)	−2.6%	—	−2.5%
Avg. Recovery after Disasters (C & GDP)	0.4	—	0.5

^{†,‡}Posterior means.

[§]From the data used in this study.

[‡]C case only.

(2011a), the empirical probability for a C (GDP) disaster per year is 3.8%, and the average disaster size is 21.5% (20.4%) for C (GDP). “The mean duration of the disasters were also similar: 3.6 years for C and 3.5 years for GDP.”

The probability of a world event (i.e., the parameter p_W) is estimated at 5.86%. Figure

1.2 depicts the estimated world event probability in each year between 1850 and 2009. Notice that the estimated world event probability is higher than 1/2 for seven year: 1914, 1917, 1930, 1931, 1940, 1944, and 2009, which are identified as the World War I (WWI), Great Depression, World War II (WWII), and the recent “Great Recession,” accordingly.

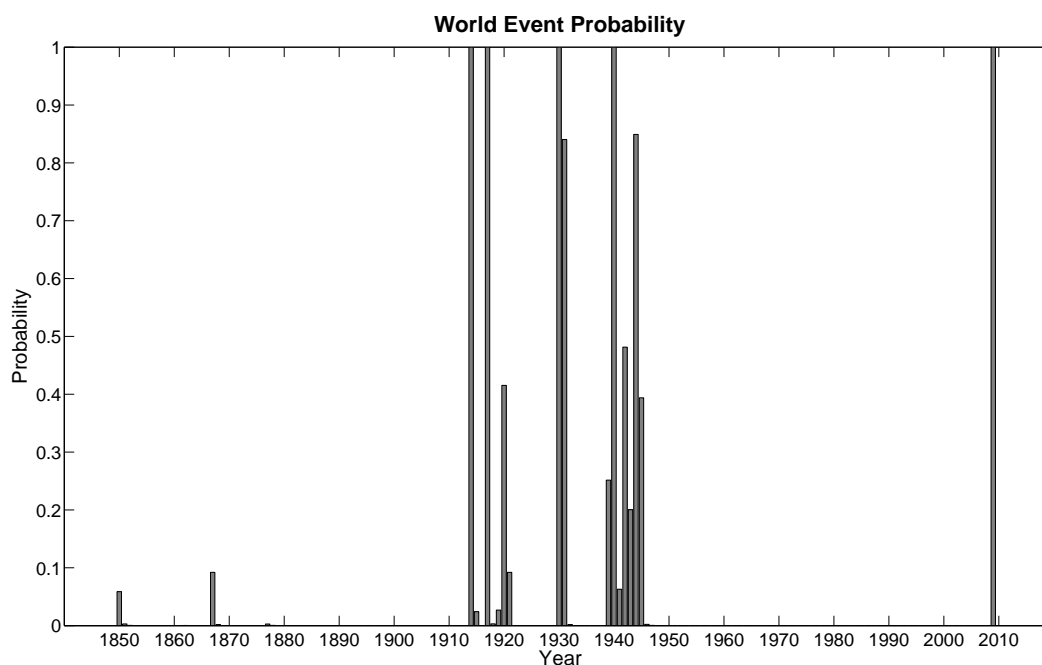


Figure 1.2: Probability of World Events

Notes: This figure plots the posterior mean of $I_{W,t}$, i.e., the estimated probability that a world event happens in each year during 1850–2009.

As mentioned previously, there is a missing data issue for the first year of the C (GDP) series of each economy. Similarly, for 2009, the latest year, there is also a problem of missing data, which renders the corresponding estimates less accurate than those for the middle years. The recent “Great Recession” is a significant event, but the corresponding “event probability” in 2009 might be overestimated because of the missing data issue. The same situation occurs in the estimation of event probability for 2009 for many countries.

The estimated probability of a world event, 5.86% per year, is significantly higher than the estimate 3.7% per year in NSBU, and there are two reasons for this difference. First, the simulation in this study indicates a world event in 2009 that is out of the scope of NSBU.⁸ Thus, the world event probability is estimated higher here. Second, the NSBU model pins down three main world event years (i.e., 1914, 1930, and 1940), while Figure 1.2 shows more “eventful years” as mentioned above. Except for 2009, all other eventful years correspond to the three disastrous events. As this study’s data set includes more countries hit by those events in different years, more years are identified as “eventful” in this study than in NSBU, which also makes the estimated world event probability higher. Other findings about rare-event/disaster process reported in Table 1.2 are discussed below in Section 1.3.2.

Rare-Event Parameters

Table 1.3 reports mean and standard deviation of the prior and posterior distributions of both the rare-event and long-run-risk parameters. The probability for a country to enter a rare event “on its own” (i.e., the parameter p_{bl}) is estimated at 1.94% per year. On average, countries have an estimated probability of 0.568 to enter events conditional on the occurrence of a world event (i.e., the parameter p_{bW}), and the total probability of a country entering a rare event is estimated to be 5.13% per year.⁹ This value is much higher than the corresponding estimate of 2.8% in NSBU.

⁸The data used in NSBU end at year 2006.

⁹See Footnote 21 of NSBU for a discussion on calculation of this probability. Note that if I simply plug in estimated parameters into the formula $p_W p_{bW} + (1 - p_W) p_{bl}$ to calculate the probability, the resulted value is 5.15% which is almost the same as the one calculated from the rigorous treatment.

Table 1.3: Rare-Event and Long-Run-Risk Parameters in the RE+LRR Model

Parameter	Prior Dist.	Prior Mean	Prior S.D.	Post. Mean	Post. S.D.
p_W	Uniform	0.0500	0.0289	0.0586	0.0179
p_{bW}	Uniform	0.500	0.289	0.568	0.0595
p_{bI}	Uniform	0.0150	0.00866	0.0194	0.00319
p_e	Uniform	0.550	0.260	0.314	0.0445
$\rho_{z,0,C}$	Uniform	0.450	0.260	0.504	0.0328
$\rho_{z,1,C}$	Uniform	1.000	0.577	0.450	0.0410
$\rho_{z,0,GDP}$	Uniform	0.450	0.260	0.577	0.0289
$\rho_{z,1,GDP}$	Uniform	1.000	0.577	0.537	0.0391
ϕ_C	Non-uniform [†]	-0.177	0.0642	-0.0850	0.00756
ϕ_{GDP}	Non-uniform [†]	-0.177	0.0642	-0.0869	0.00528
θ_C	Normal	0.000	0.400	-0.0262	0.00841
θ_{GDP}	Normal	0.000	0.400	-0.0260	0.00851
$\sigma_{\phi,C}$	Non-uniform [†]	0.0987	0.0475	0.0628	0.00569
$\sigma_{\phi,GDP}$	Non-uniform [†]	0.0987	0.0475	0.0651	0.00398
$\sigma_{\theta,C}$	Uniform	0.130	0.0693	0.152	0.0130
$\sigma_{\theta,GDP}$	Uniform	0.130	0.0693	0.157	0.0152
λ_C	Uniform	0.255	0.141	0.428	0.0441
λ_{GDP}	Uniform	0.255	0.141	0.295	0.0546
w_C	Uniform	50.0	28.9	2.64	0.286
w_{GDP}	Uniform	50.0	28.9	2.45	0.252

[†]The prior distributions of ϕ_*^\diamond and $\sigma_{\phi,*}^\diamond$ are uniform, and correspondingly, the prior distributions of ϕ_* and $\sigma_{\phi,*}$ are non-uniform.

To assess the difference in the two estimates of a country's total event probability, I examine a special type of events—rare disasters—as well as rare events. The *ex ante* changes of $\log C$ (GDP) due to a rare event relative to their original trend are depicted in Figure 1.3 (1.4). These figures illustrate the *ex ante* event distributions from the perspective of agents who solely know that no event occurs in year 0 and a rare event begins in year 1. (The size and the duration of the event are unknown to the agents.) These figures show that the influences of rare events on the economy are dispersed, and some events are clearly no disasters. The estimation reveals that in *ex ante*, 64.5% (65.5%) of the rare events are C (GDP) disasters that cause “peak-to-trough” contractions in C (GDP) of about 10% or more. The actual threshold is taken to be a decrease of 10 log points, which equals 9.5% ($= 1 - e^{-0.1}$), the value taken in Barro and Jin (2011a). For C (GDP) disasters, the *ex ante* average “peak-to-trough” contraction in C (GDP) is 23%, and the average duration is 3.2 (3.3) years. For a given country, the total probability of entering a disaster is estimated at 3.31% ($= 5.13\% \times 0.645$) per year for C and 3.36% ($= 5.13\% \times 0.655$) per year for GDP. The total disaster probability for a country is not reported in NSBU. However, as not every event is a disaster, the total disaster probability must be less than 2.8%.¹⁰ Similar to Figure 1.3 (1.4), Figure 1.5 (1.6) depict the *ex ante* changes of $\log C$ (GDP) due to a disaster relative to their original trends.

The estimate of p_e , the probability of a country exiting an event, is 0.314, which corresponds to an average event duration of 3.2 years. NSBU's estimate of p_e is 0.165, which corresponds to a longer disaster duration (roughly 6 years). For disasters, the average durations are almost the same: 3.2 years for C , and 3.3 years for GDP.

¹⁰Note that the term “rare disasters” used in NSBU corresponds to the term “rare events” in this study.

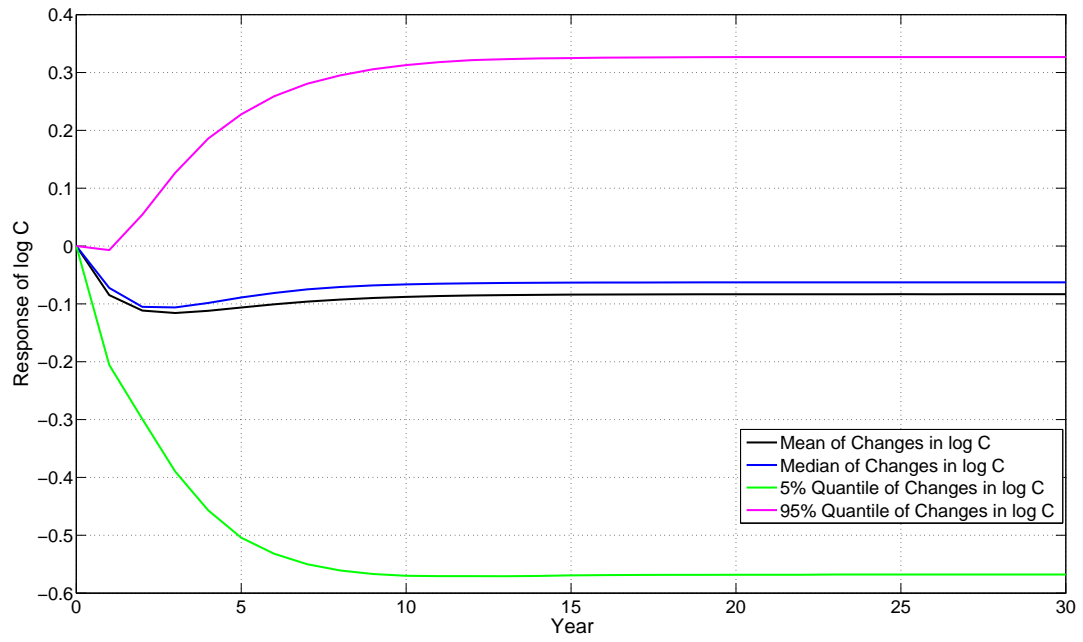


Figure 1.3: Ex Ante Rare-Event Distribution (C)

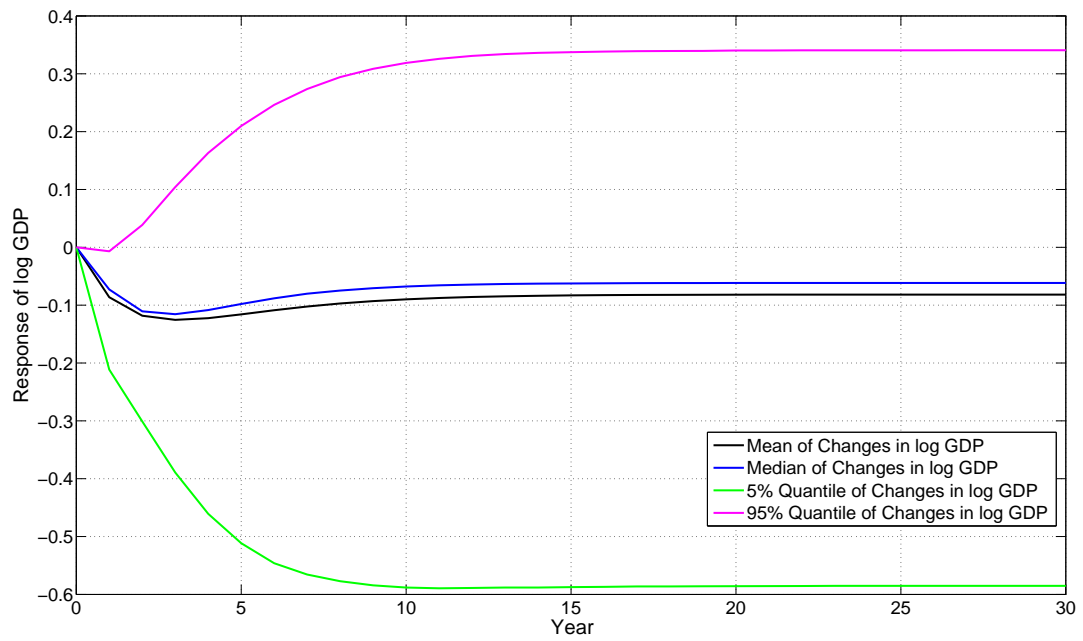


Figure 1.4: Ex Ante Rare-Event Distribution (GDP)

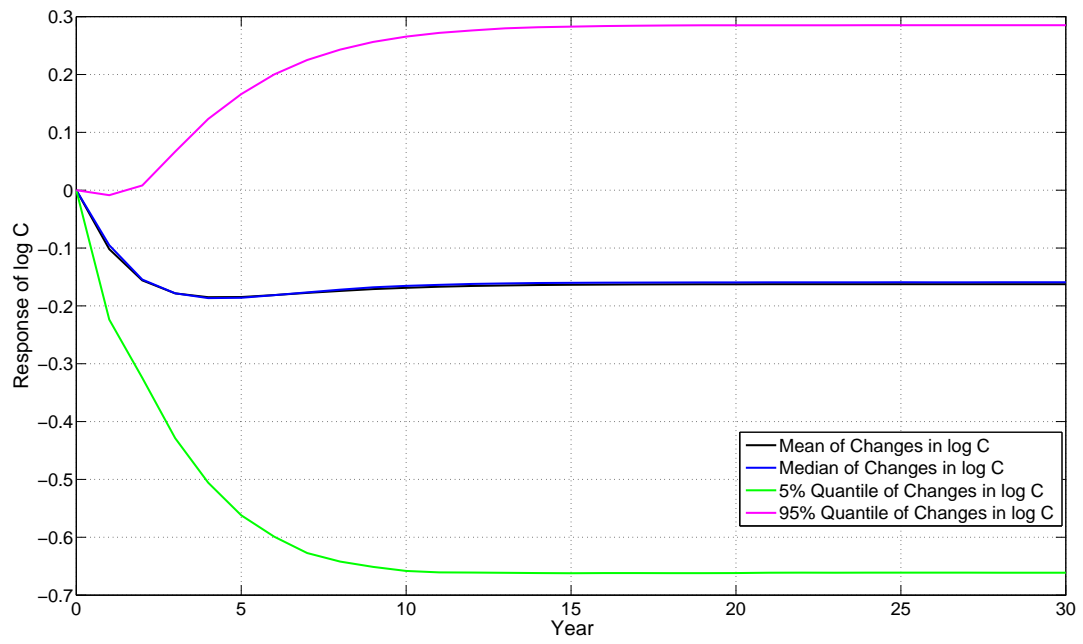


Figure 1.5: Ex Ante Disaster Distribution (C)

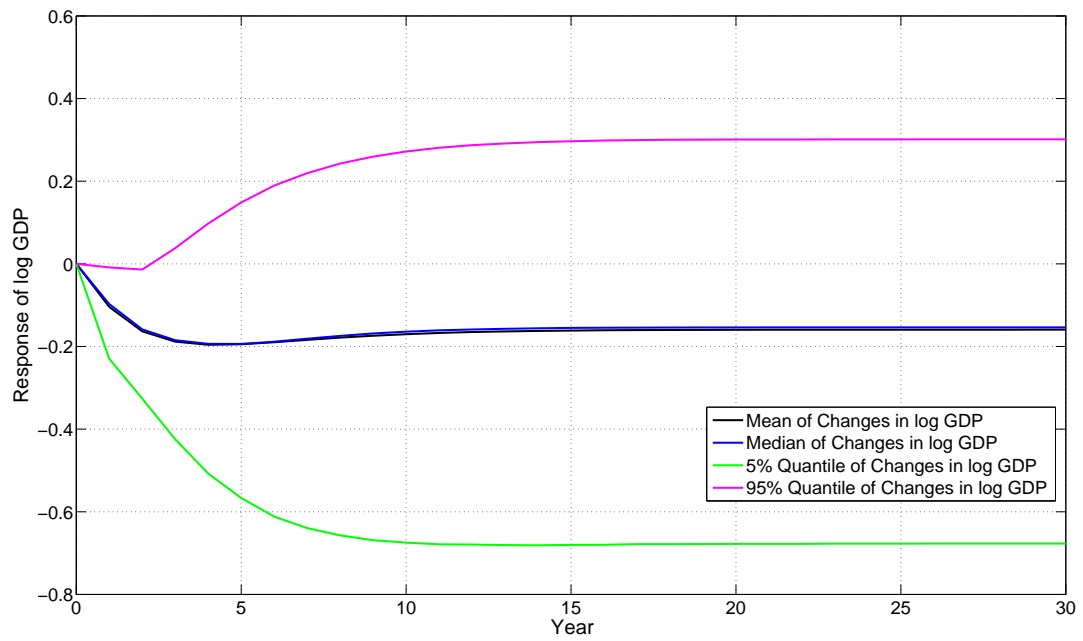


Figure 1.6: Ex Ante Disaster Distribution (GDP)

In this study the estimated short-run and permanent event effects are given as $\hat{\phi}_C = -0.0850$, $\hat{\phi}_{GDP} = -0.0869$, $\hat{\theta}_C = -0.0262$, and $\hat{\theta}_{GDP} = -0.0260$. Thus, in an average event, the short-run shock is to reduce C (GDP) by 8.2% (8.3%) per year, while the permanent impact is -2.6% per year. This result is slightly different from that of NSBU, where $\hat{\phi}_C = -0.111$, $\hat{\theta}_C = -0.025$. Relative to NSBU's estimates, (1) the estimated average short-run shock of events is smaller, and (2) the average duration of events is shorter, but (3) events are more frequent.

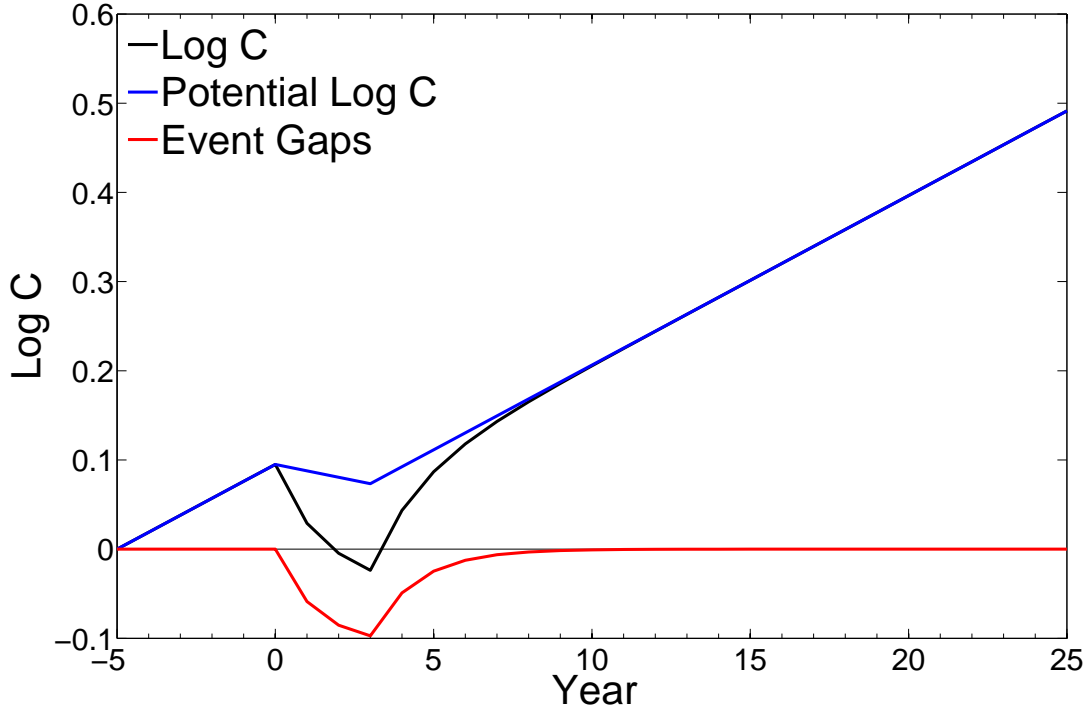


Figure 1.7: Response of log C to a Typical Rare Event

Compared with NSBU, the estimates of the standard deviations of the short-run shocks, $\hat{\sigma}_{\phi,C} = 0.0628$ and $\hat{\sigma}_{\phi,GDP} = 0.0651$, are a bit smaller than the value in NSBU ($\hat{\sigma}_{\phi,C} = 0.083$). Together with the estimates of ϕ_* , the estimates in this study indicate smaller short-run

event shocks. On the other hand, the estimates of the standard deviations of the permanent effects, $\hat{\theta}_{\phi,C} = 0.152$ and $\hat{\theta}_{\phi,GDP} = 0.157$, are larger than the value in NSBU ($\hat{\theta}_{\phi,C} = 0.121$). Together with the estimates of θ_* , the permanent event effects are estimated to be “riskier” here than in NSBU. Note that the magnitude of $\hat{\sigma}_{\theta,*}$ is about five times larger than that of $\hat{\theta}_*$, which means for 43% periods of events, the permanent event effects are positive. Some of these events actually correspond to the “occasional bonanzas” mentioned in Barro and Jin (2011a). One typical example of bonanzas (or booms) is the GDP case of the United States in the WWII, which is depicted in Figure 1.9. In NSBU, the estimates are similar: the permanent event effects are positive for 42% event periods.

To visualize the effects of the rare-event parameters estimated above, I draw a graph of the response of $\log C$ to a “typical event” in Figure 1.7. The black curve is the observed $\log C$, the blue curve is the potential level of $\log C$, and the red curve represents the event gap. The typical event lasts for three years from Year 1 to Year 3, and the short-run and long-run event effects are set to be the estimated values as mentioned above. Note that the graph in Figure 1.7 is for a typical event. An actual event can be a much worse disaster, or a much better boom.

As a summary, the rare event/disaster process estimates accord well with the facts observed in data (as documented in Section 1.3.2), and are also close to the findings in Barro and Jin (2011a).

1.3.3 Empirical Findings on Long-Run Risks

In this section, I discuss empirical findings on the long-run growth rates characterized by parameters λ_* and w_* , and findings on the estimated long-run growth rates for each

economy.

Long-Run-Risk Parameters

The estimated Poisson parameter λ_C (λ_{GDP}) is 0.428 (0.295). In other words, the probability that a country's long-run growth rates remain unchanged in a year is 65.2% (74.5%) for C (GDP), which indicates that the long-run growth rates are relatively stable. The estimated weight parameter $\hat{w}_C = 2.64$ and $\hat{w}_{GDP} = 2.45$, which staunchly supports the economic intuition that the long-run growth rates are mean-reverting.

It is important to compare my estimates of long-run risks to the BY's and BKY's parameter values. For that end, I convert my estimates into a monthly autoregressive coefficient ρ (corresponding to ρ in Equation (1.12)) and the standard deviation σ of innovations (corresponding to $\phi_e\sigma$ in Equation (1.12)). The criterion for this conversion is to match up the autocovariance functions (ACF's). (In Appendix A.4, I explain the details of this conversion.)

Table 1.4 displays some key parameters in BY's and BKY's calibration and the corresponding (converted) parameter estimated from the proposed model. The table shows that the long-run growth rates $\{\mu_{i,t,*}\}_t$ estimated by the model are very persistent; the corresponding monthly autoregressive coefficient ρ is 0.988, larger than the calibrated value 0.979 in BY and 0.975 in BKY. The converted σ is 0.0018, 4–5 times larger than BY's and BKY's value of $\phi_e\sigma$ —the standard deviation of the innovation in the persistent predictable component of the consumption growth, as described by Equation (1.12). Moreover, the value 0.0018 of σ is more than 70% larger than BY's and BKY's value of $\phi\phi_e\sigma$ —the standard deviation of the innovation in the persistent predictable component of the dividend

growth. Thus, compared with BY's and BKY's calibration, this study does not underestimate the long-run-risk component ($\triangleq \mu_{i,t,*} - \beta_i$).

Table 1.4: Comparison to BY's and BKY's Parameter Values

Parameter in BY/BKY	Symb.	BY	BKY	Symb.	Model
Mean Growth (C)	μ	0.0015	0.0015	β_i	0.019 [†]
LRR Persistence	ρ	0.979	0.975	ρ	0.988
LRR Volatility Multiple	φ_e	0.044	0.038		
Mean Growth (Dividend)	μ_d	0.0015	0.0015		
Dividend Leverage	ϕ	3.0	2.5		
Dividend Volatility Multiple	φ_d	4.50	5.96		
Dividend Consumption Exposure	π	0.0	2.6		
Baseline Volatility	σ	0.0078	0.0072	$\sigma_{\eta,i,*}$	0.021 [†]
	$\varphi_e \sigma$	0.000343	0.000274	σ	0.0018
	$\phi \varphi_e \sigma$	0.00103	0.000684		
	$\varphi_d \sigma$	0.0351	0.0429		
	$\pi \sigma$	0.000	0.0187		
				λ_*	0.36
				w_*	2.5

[†]Annual values.

Long-Run Growth Rates

One important “by-product” of the empirical model is the estimated long-run growth rates for the 42 economies. As rare events and long-run risks are intertwined in the data, the estimation of the long-run growth rates is not feasible without an explicit model of rare events. The empirical model here works as a “filter” and tells us each economy’s growth trend in a long period of time. For instance, the red curves in Figure 1.8 and 1.9 demonstrate the long-run growth rates of C and GDP for the United States during 1850–2009, respectively. Investigation of those curves can be relevant for understanding economic growth, development and history.

1.3.4 Distinctions between Rare Events and Long-Run Risks

Unlike the claim that “cyclical risks” contain disaster risks in Bansal, Kiku and Yaron (2010), the empirical results on the decomposition of growth gaps, as defined in Section 1.2.3, indicate that rare events and long-run risks are distinct risks. Figures 1.10 and 1.11 depict the decomposition of demeaned growth gaps for C and GDP of the United States. These figures illustrate the distinct features of the rare-event and long-run-risk components, and I summarize them as “sporadic, drastic, and jumping outbursts” and “persistent, moderate, and smooth fluctuations,” respectively.

Table 1.5 lists the mean, volatility, and the first-order autocorrelation of the rare-event, long-run-risk, and noise components, and the demeaned growth gaps for C and GDP cases of the United States and the world. The table indicates that the standard deviations of rare-event and noise components are more than one times larger than those of the long-run-risk components.

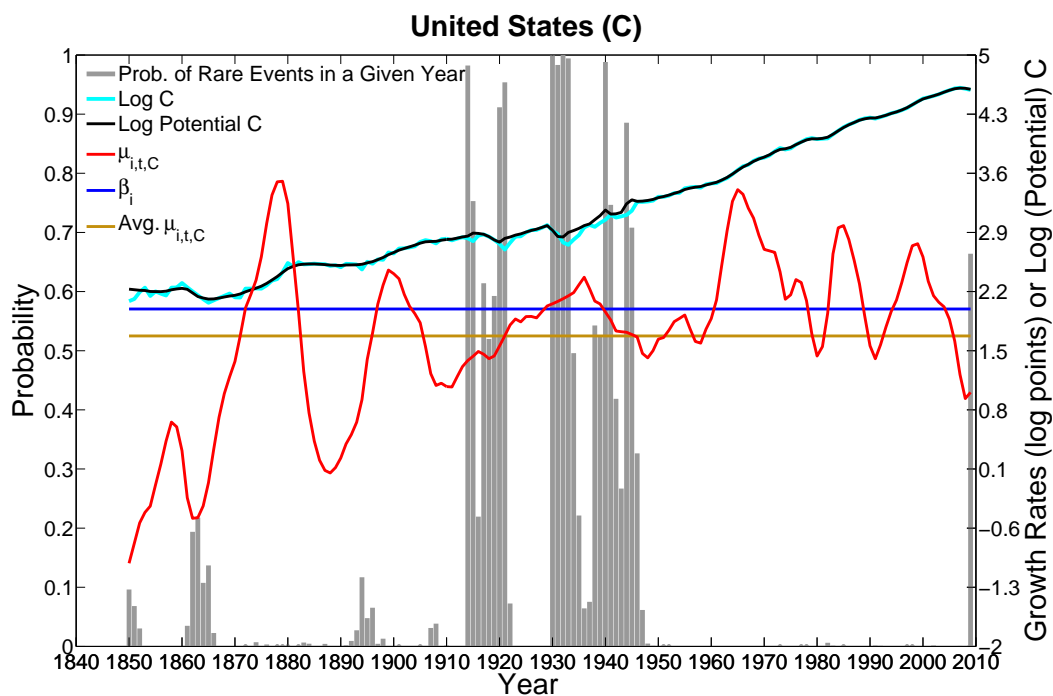


Figure 1.8: Estimates for the United States (C)

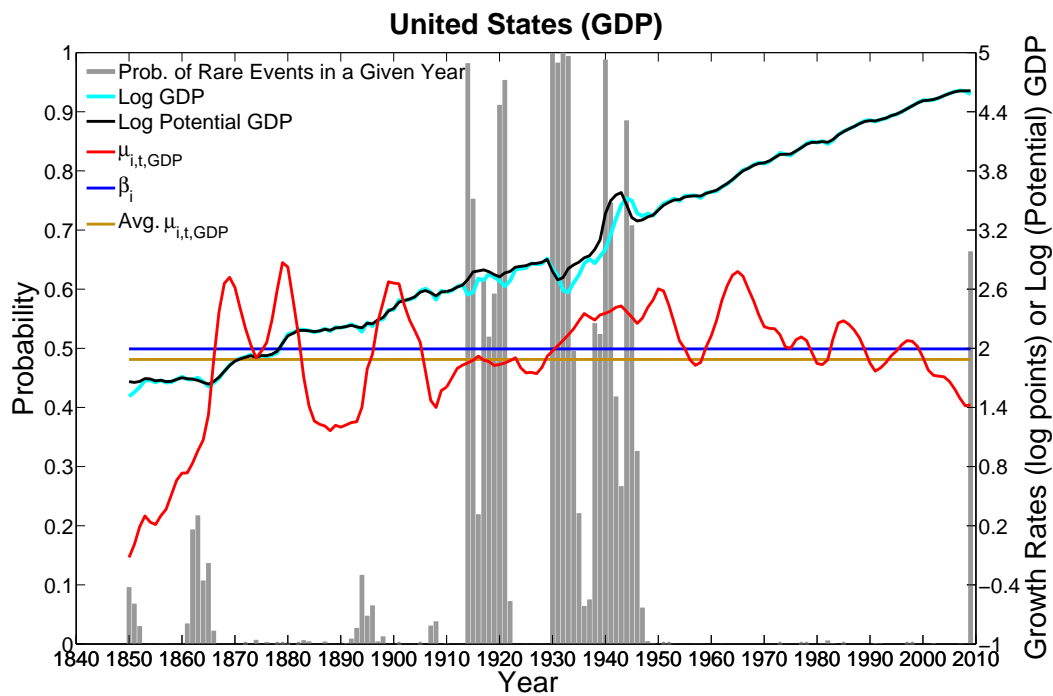


Figure 1.9: Estimates for the United States (GDP)

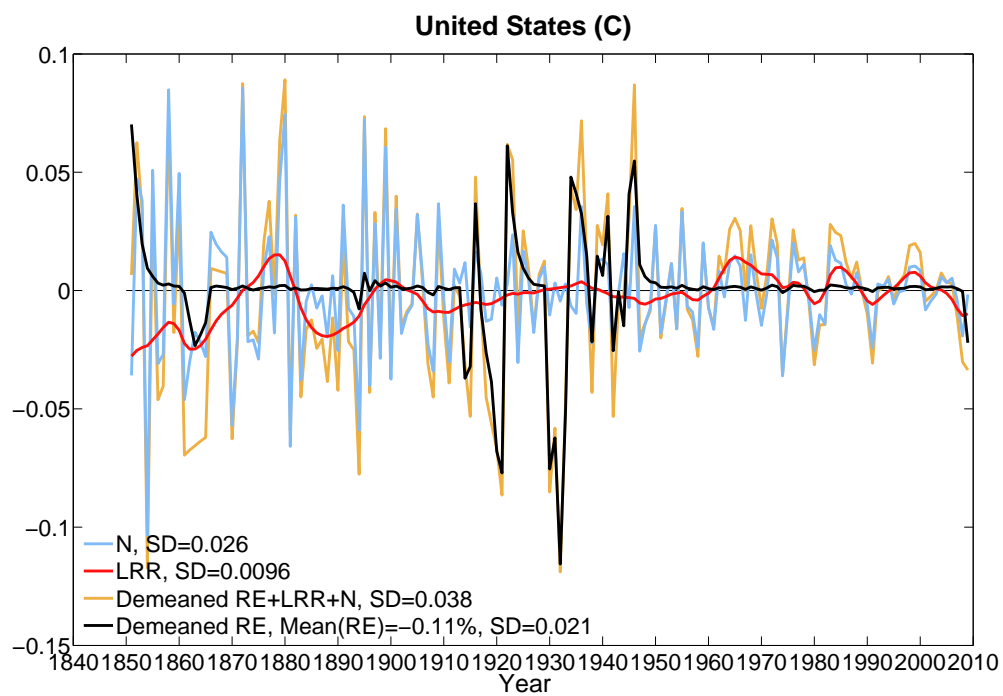


Figure 1.10: Demeaned Growth Gap Decomposition for United States (C)

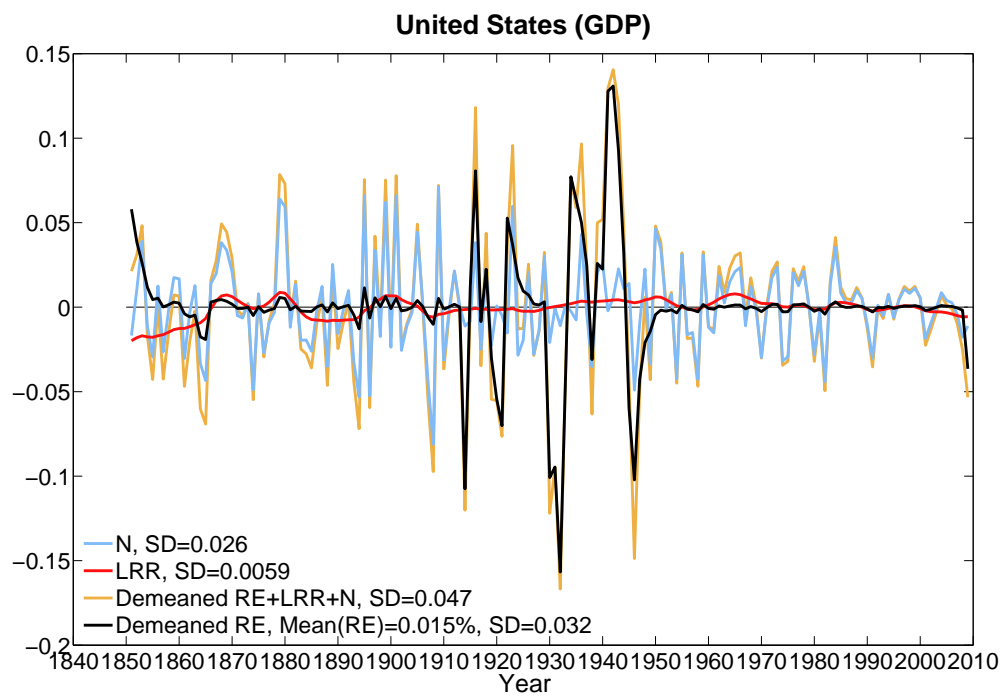


Figure 1.11: Demeaned Growth Gap Decomposition for United States (GDP)

Table 1.5: The Mean, Standard Deviation, and First-order Autocorrelation of the RE, LRR, and Noise Components, and the Demeaned Growth Gaps for C and GDP of the US and the World

	U.S.				World Avg.			
	$RE_{i,t,*}$	$LRR_{i,t,*}$	$N_{i,t,*}$	$\Delta y_{i,t,*}^{DM}$	$RE_{i,t,*}$	$LRR_{i,t,*}$	$N_{i,t,*}$	$\Delta y_{i,t,*}^{DM}$
(C)								
Mean	-0.106				-0.270			
σ	2.09	0.963	2.63	3.76	3.88	1.17	3.82	6.15
AC1	0.482	0.972	-0.369	0.0817	0.252	0.980	-0.144	0.133
(GDP)								
Mean	0.0146				-0.254			
σ	3.19	0.587	2.63	4.71	3.88	1.18	2.84	5.72
AC1	0.547	0.971	-0.201	0.265	0.410	0.986	-0.241	0.179

Table 1.6 displays the correlation coefficients between the rare-event, long-run-risk, and noise components, and the demeaned growth gaps for C and GDP of the United States and the world. It is interesting to note that any two of them are positively correlated except the pair of the rare-event and long-run-risk components. The rare-event and long-run-risk components are basically uncorrelated, and the most salient positive correlations are between the rare-event component and the demeaned growth gap, and between the noise component and the demeaned growth gap.

Table 1.6: Correlation Coefficients between the RE, LRR, and Noise Components, and the De-meaned Growth Gaps for C and GDP of the US and the World

(a) C and GDP cases for the United States

	$RE_{i,t,C}$	$LRR_{i,t,C}$	$N_{i,t,C}$	$\Delta y_{i,t,C}^{DM}$	
$RE_{i,t,GDP}$		-0.0401	0.115	0.622	$RE_{i,t,C}$
$LRR_{i,t,GDP}$	-0.00844		0.172	0.354	$LRR_{i,t,C}$
$N_{i,t,GDP}$	0.238	0.262		0.807	$N_{i,t,C}$
$\Delta y_{i,t,GDP}^{DM}$	0.806	0.265	0.752		$\Delta y_{i,t,C}^{DM}$
	$RE_{i,t,GDP}$	$LRR_{i,t,GDP}$	$N_{i,t,GDP}$	$\Delta y_{i,t,GDP}^{DM}$	

(b) C and GDP cases for the World

	$RE_{i,t,C}$	$LRR_{i,t,C}$	$N_{i,t,C}$	$\Delta y_{i,t,C}^{DM}$	
$RE_{i,t,GDP}$		0.00173	0.135	0.691	$RE_{i,t,C}$
$LRR_{i,t,GDP}$	0.00165		0.164	0.303	$LRR_{i,t,C}$
$N_{i,t,GDP}$	0.259	0.157		0.710	$N_{i,t,C}$
$\Delta y_{i,t,GDP}^{DM}$	0.790	0.292	0.715		$\Delta y_{i,t,C}^{DM}$
	$RE_{i,t,GDP}$	$LRR_{i,t,GDP}$	$N_{i,t,GDP}$	$\Delta y_{i,t,GDP}^{DM}$	

The fundamental distinctions between rare events and long-run risks are as follows. First, long-run risks are persistent, while rare events are not. Many rare macroeconomic events burst out suddenly and unexpectedly, causing drastic changes (mostly declines) in consumption and output. Previous studies show that most of the observed macroeconomic disasters happened in periods of world disasters, such as World Wars I and II,

the Great Influenza, and the Great Depression, and in periods of idiosyncratic disasters, such as regional wars, coups, or revolutions. Figures 1.10 and 1.11 visualize the rare-event component's sporadic outbursts—oscillating sharply during event periods and diminishing quickly afterwards—and the long-run-risk component's persistent and smooth fluctuations.

Second, the volatilities of the rare-event and long-run-risk components are different. From the moments displayed in Table 1.5 and the decomposition of demeaned growth gaps illustrated in Figures 1.10 and 1.11, we see that the volatilities of the rare-event and noise components are significantly larger than the volatility of the long-run-risk component.

Third, the rare-event and long-run-risk components have different durations. In theory, the movement of the long-run-risk component is random and not periodic. However, the empirical results indicate that the long-run growth rates fluctuate up and down like cycles, which I call “long-run growth cycles.” The model estimates show that durations of rare events are much shorter than those of long-run growth cycles. The average durations of C and GDP disasters are 3.5 years in the data, and are estimated at 3.2 and 3.3 years, respectively. On the contrary, the long-run growth cycles endure much longer. (Figures 1.10 and 1.11 visualize this.)

1.3.5 Comparison of Models

Based on the proposed RE+LRR model, we obtain an RE model by shutting down the long-run-risk channel, i.e., by setting $\lambda_* = 0$. Similarly, we obtain an LRR model by shutting down the rare-event channel, i.e., by setting the rare-event probabilities (i.e., p_W and p_{bl}) to be 0. Thus, the proposed RE+LRR model contains the RE and LRR models as special cases.

As rare events and long-run risks coexist in asset markets, the RE+LRR model should be the appropriate one among the three models studied in this paper. Furthermore, as the goal of this study is to identify and evaluate these two types of risks, the RE+LRR model is the right choice. Therefore, we do not need to do model selection here. However, it is still informative to estimate the RE and LRR models using the data and compare the parameter estimates.

The estimation procedure for the RE+LRR model is discussed in Section 1.3.1. For RE and LRR models, I run four simulation chains for 200,000 iterations each using the same data, and the first 100,000 iterations are dropped as burn-in. The estimates of parameters of the RE and LRR models are displayed in Table 1.7 and 1.8, respectively. When comparing Table 1.7 with Table 1.3, we see that the RE model provides larger estimates for event probabilities (i.e., p_W , p_{bW} , and p_{bI}), average event duration (characterized by p_e^{-1}) and event gap persistence (i.e., $\rho_{z,j,*}$), but the estimated short-run and permanent event effects (i.e., ϕ_* and θ_* , respectively) are smaller. When comparing Table 1.8 with Table 1.3, we observe that the long-run-risk component estimated by the LRR model is more volatile. In comparison with the RE+LRR model, the changes in the estimates of parameters of the RE and LRR models are plausible and in accord with the economic intuition.

Table 1.7: Rare-Event Parameters in the RE Model

Parameter	Prior Dist.	Prior Mean	Prior S.D.	Post. Mean	Post. S.D.
p_W	Uniform	0.0500	0.0289	0.0617	0.0180
p_{bW}	Uniform	0.500	0.289	0.594	0.0564
p_{bI}	Uniform	0.0150	0.00866	0.0244	0.0032
p_e	Uniform	0.550	0.260	0.260	0.0296
$\rho_{z,0,C}$	Uniform	0.450	0.260	0.578	0.0466
$\rho_{z,1,C}$	Uniform	1.000	0.577	0.492	0.0420
$\rho_{z,0,GDP}$	Uniform	0.450	0.260	0.605	0.0394
$\rho_{z,1,GDP}$	Uniform	1.000	0.577	0.545	0.0395
ϕ_C	Non-uniform [†]	−0.177	0.0642	−0.0732	0.0066
ϕ_{GDP}	Non-uniform [†]	−0.177	0.0642	−0.0737	0.0064
θ_C	Normal	0.000	0.400	−0.0100	0.0059
θ_{GDP}	Normal	0.000	0.400	−0.0093	0.0055
$\sigma_{\phi,C}$	Non-uniform [†]	0.0987	0.0475	0.0544	0.0051
$\sigma_{\phi,GDP}$	Non-uniform [†]	0.0987	0.0475	0.0554	0.0049
$\sigma_{\theta,C}$	Uniform	0.130	0.0693	0.140	0.0094
$\sigma_{\theta,GDP}$	Uniform	0.130	0.0693	0.142	0.0107

[†]See Notes of Table 1.3.

Table 1.8: Long-Run-Risk Parameters in the LRR Model

Parameter	Prior Dist.	Prior Mean	Prior S.D.	Post. Mean	Post. S.D.
λ_C	Uniform	0.255	0.141	0.495	0.00540
λ_{GDP}	Uniform	0.255	0.141	0.495	0.00453
w_C	Uniform	50.0	28.9	2.12	0.151
w_{GDP}	Uniform	50.0	28.9	2.00	0.117

1.4 Evaluation in Asset Pricing

In this section, I first discuss the asset pricing condition and the values of IES ψ , CRRA γ and (subjective) discount factor (DF) β . Then I evaluate the contributions of rare events and long-run risks to various asset pricing statistics, including the equity premium, the Sharpe ratio, the mean, first-order autocorrelation, and volatility of the log price-dividend ratio (PDR) on equity (the market), and volatilities of the risk-free rate, market return, and equity premium.

1.4.1 Values of IES ψ , CRRA γ , and DF β

The asset pricing implications of the estimated model are analyzed following Mehra and Prescott (1985) and other studies. To detach the connection between the CRRA and the IES, it is assumed that the representative agents have the EZW preferences. For this type of preferences, Epstein and Zin (1989) work out the condition that the return on any asset

needs to satisfy—namely, the return is given by the solution to the following equation

$$E_t \left[\beta^{\bar{\zeta}} \left(\frac{C_{i,t+1}}{C_{i,t}} \right)^{-\bar{\zeta}/\psi} R_{w,i,t+1}^{-(1-\bar{\zeta})} R_{a,i,t+1} \right] = 1, \quad (1.15)$$

where $R_{a,i,t+1}$ is the gross return on a given asset a in country i from period t to $t + 1$, $R_{w,i,t+1}$ is the gross return on wealth of the agent in country i , which equals the C (GDP) stream described in the model, β is the DF, ψ is the IES, and $\bar{\zeta} \triangleq \frac{1-\gamma}{1-1/\psi}$ (γ is the CRRA).

Since the proposed model with EZW preferences cannot be solved in closed form, I adopt a numerical method. Different from solving a loglinear approximate model proposed by Campbell and Shiller (1988) and Campbell (1993), here I directly solve the nonlinear integral equation on a grid. Specifically, Equation (1.15) gives a recursive formula for the PDR of the consumption (output) claim, and the iteration converges to the fixed point of that PDR. Then the fixed point of the PDR for any other asset can be solved through Equation (1.15).

In order to analyze the the asset pricing implications of the empirical model, we need the parameters estimated in Section 1.3, as well as the values of IES ψ , CRRA γ , and DF β . There has been a debate about the value of IES in macroeconomics and finance literature: Hall (1988) estimates the IES to be close to zero, Campbell (2003) and Guvenen (2009) claim that it should be less than 1, Tsai and Wachter (2012) assumes IES to be 1, BY use a value of 1.5, and Barro (2009) adopts the analysis in Gruber (2006) and take IES to be 2. NSBU provides evidences that low IES values are inconsistent with the observed behavior of asset prices during consumption disasters. As noted in BY, in order to have an increase in the wealth-to-consumption ratio in response to higher expected rates of return, IES ψ must be greater than 1, where the intertemporal substitution effect dominates the wealth

effect. Barro (2009) also requires that $IES > 1$ to avoid the counterfactual prediction that an increase in uncertainty implies a higher price-dividend ratio. For these reasons and to compare results, I follow Gruber (2006), Barro (2009), and NSBU and take $\psi = 2$. I will show how the results change with variations in the value of IES in Section 1.4.2.

To determine the values of CRRA γ and DF β , I consider two assets—the one period risk-free bill and an unleveraged claim on the consumption (output) process described in the empirical model. I choose γ and β to simultaneously match the risk-free rate and the unlevered market return, or equivalently, to match the risk-free rate and the equity premium. Other asset pricing statistics are calculated after the values of γ and β are determined.

The target values of asset pricing statistics come from two sources. For the risk-free rate and market return, I adopt the values documented in Barro and Ursúa (2008): the average arithmetic real rate of return on short-term bills is 0.9% per year; on stock, 8.1% per year. Then the average equity premium is 7.2% per year. For other asset pricing statistics, I adopt the values documented in Nakamura, Sergeyev and Steinsson (2012). Some values (e.g. the market return) correspond to levered claim on the consumption (output) stream, and I adjust these values for leverage in corporate financial structure with a debt-equity ratio of 0.5. These data are displayed in Table 1.9, and unleveraged target values of the asset pricing statistics are listed in the rightmost column, where the unleveraged equity premium is 4.8%.

With the parameters estimated from the RE+LRR model, matching the risk-free rate and the unlevered market return gives $\gamma = 6.3$ and $\beta = \exp(-0.031)(\approx 0.97)$. This suggests that there is still room for other factors, such as habit (Campbell and Cochrane, 1999), to play roles in generating the equity premium.

Table 1.9: Asset Pricing Statistics: Data[†]

Moment	Data1 ¹	U1 ²	Data2 ³	U2 ⁴	Data3 ⁵
$E(R_f)$	1.0113		1.009		1.009
$E(R_m)$	1.0823	1.0586	1.081	1.057	1.057
$E(R_m - R_f)$	0.0710	0.0473	0.072	0.048	0.048
$\sigma(R_f)$	0.0333				0.0333
$\sigma(R_m)$	0.1789	0.1225			0.1225
$\sigma(R_m - R_f)$	0.1737	0.1158			0.1158
$\frac{E(R_m - R_f)}{\sigma(R_m - R_f)}$	0.41	0.41			0.41
$E(p - d)$	3.30				3.30
$\sigma(p - d)$	0.40				0.40
$AC1(p - d)$	0.90				0.90

[†]The expressions $E(R_f)$, $E(R_m)$, and $E(R_m - R_f)$ are average annual risk-free rate, market return, and equity premium, respectively; the expressions $\sigma(R_f)$, $\sigma(R_m)$, and $\sigma(R_m - R_f)$ are volatilities of the risk-free rate, market return, and equity premium, respectively; the expression $\frac{E(R_m - R_f)}{\sigma(R_m - R_f)}$ is the Sharpe ratio; the expressions $E(p - d)$, $\sigma(p - d)$, and $AC1(p - d)$ are the mean, volatility, and first-order autocorrelation of the log PDR on equity, respectively.

¹U.S. data, source: Nakamura, Sergeyev and Steinsson (2012).

²Unleveraged values, obtained by applying debt-equity ratio of 0.5 to Data1.

³Average of 17 OECD economies, source: Barro and Ursúa (2008).

⁴Unleveraged values, obtained by applying debt-equity ratio of 0.5 to Data2.

⁵Target values from the left four columns.

1.4.2 Empirical Evaluation

In this subsection, I first discuss the asset pricing statistics calculated by using parameters that match the risk-free rate and the market return. Then I explore how the variations in the values of CRRA γ and IES ψ affect the asset pricing statistics. Finally, I compare the asset pricing statistics calculated by using parameter values reported in Table 1.3, 1.7, and 1.8.

Asset Pricing Statistics

Table 1.10 displays the target values and model predicts of the asset pricing statistics. The “Model 0” column are calculated using parameters estimated in the RE+LRR model together with $\psi = 2$, $\gamma = 6.3$, and $\beta = \exp(-0.031)$. The “Model 1” column are calculated using the same parameter values as in Model 0, except for setting $\lambda = 0$. Namely, the long-run-risk channel is shut down in Model 1, the RE model. The “Model 2” column are calculated using the same parameter values as in Model 0, except for setting the rare-event probabilities to be 0. Namely, the rare-event channel is shut down in Model 2, the LRR model.

The equity premium. Table 1.10 indicates that the equity premium in Model 0 (4.8%) is not far from the sum (5.2%) of the corresponding values in Model 1 (4.2%) and Model 2 (1.0%). The effects of rare events and long-run risks basically add to each other in generating the equity premium, which is not the case for other asset pricing statistics. The reasons for this phenomenon lie in the features of and interrelations between rare events and long-run risks as discussed in Section 1.3, and in the functional form of Equation (1.15). Table 1.10 demonstrates that this is also the case for other values of CRRA γ . These results

Table 1.10: Asset Pricing Statistics: Data and Model Values for Various CRRA γ [†]

Moment	Data ³	Model 0 ² (RE+LRR)	Model 1 ³ (RE)	Model 2 ⁴ (LRR)	Model 0a ⁵ ($\gamma = 7$)	Model 1a ⁵ ($\gamma = 7$)	Model 2a ⁵ ($\gamma = 7$)	Model 0b ⁶ ($\gamma = 10$)	Model 1b ⁶ ($\gamma = 10$)	Model 2b ⁶ ($\gamma = 10$)
$E(R_f)$	1.009	1.009	1.012	1.034	1.000	1.004	1.033	0.954	0.958	1.027
$E(R_m)$	1.057	1.057	1.054	1.044	1.060	1.058	1.045	1.078	1.074	1.048
$E(R_m - R_f)$	0.048	0.048	0.042	0.010	0.061	0.054	0.012	0.124	0.116	0.021
$\sigma(R_f)$	0.0333	0.0267	0.0262	0.0302	0.0263	0.0258	0.0302	0.0250	0.0247	0.0304
$\sigma(R_m)$	0.1225	0.0683	0.0643	0.0653	0.0695	0.0658	0.0653	0.0711	0.0681	0.0658
$\sigma(R_m - R_f)$	0.1158	0.0670	0.0633	0.0691	0.0681	0.0646	0.0692	0.0675	0.0646	0.0697
$\frac{E(R_m - R_f)}{\sigma(R_m - R_f)}$	0.41	0.719	0.661	0.145	0.890	0.837	0.168	1.83	1.80	0.302
$E(p - d)$	3.30	3.26	3.32	3.65	3.17	3.23	3.63	2.83	2.88	3.52
$\sigma(p - d)$	0.40	0.0574	0.0420	0.0455	0.0575	0.0431	0.0456	0.0559	0.0443	0.0472
$AC1(p - d)$	0.90	0.627	0.406	0.800	0.619	0.413	0.801	0.594	0.423	0.806
$AC2(p - d)$		0.459	0.184	0.702	0.447	0.189	0.702	0.413	0.197	0.706

[†] $AC2(p - d)$ is the second-order autocorrelation of the log PDR on equity; see Notes of Table 1.9 for explanation on other notations.

¹Source: See Table 1.9.

²Calculated using parameters estimated from the RE+LRR model together with $\psi = 2$, $\gamma = 6.3$, and $\beta = \exp(-0.031)$.

³Calculated using the same parameter values as in Model 0, except for setting $\lambda = 0$ (shutting down LRR, i.e., the RE model).

⁴Calculated using the same parameter values as in Model 0, except for setting the RE probabilities to be 0 (shutting down RE, i.e., the LRR model).

⁵Column “Model ia ” ($i = 0, 1$, or 2) is calculated using the same parameter values as in Column “Model i ,” except for setting $\gamma = 7$.

⁶Column “Model ib ” ($i = 0, 1$, or 2) is calculated using the same parameter values as in Column “Model i ,” except for setting $\gamma = 10$.

clearly indicate that the main portion of the equity premium is the compensation for rare events, and the contribution of long-run risks to equity premium is not significant.

The risk-free rate volatility. For the volatility of the risk-free rate, the Model 0 value (0.0267) is only slightly larger than the Model 1 value (0.0262), and is smaller than the Model 2 value (0.0302), which is smaller than the target value (0.0333). Unlike long-run risks, rare events do not have persistent effects on the expected growth rate of consumption. Thus, rare events generate a relatively smaller risk-free rate volatility than do the long-run risks. Note that adding rare events to the long-run-risk model does not increase the predicted risk-free rate volatility (and even makes it smaller), which is remarkably different from the equity premium case.

The market return and equity premium volatility. For the market volatility, Model 0 gives the closest estimate (0.0683), which is 44% smaller than the target value (0.1225). For the volatility of the equity premium, Model 2 gives the closest estimate (0.0691), which is 40% smaller than the target value (0.1158). For these two volatilities, values given by Model 0, 1, and 2 are close to each other, and are smaller than the target values.

The Sharpe ratio. Relative to the target value of the Sharpe ratio (0.41), the values in Model 0, 1, and 2 are 75% higher, 61% higher, and 65% lower, respectively. Based on the relative deviation, Model 1 provides the best fit. As Model 0, 1, and 2 provide similar equity premium volatility, the differences in predicted Sharpe ratio are mainly caused by the differences in predicted equity premium.

The mean and volatility of the log PDR on equity. Both Model 0 and 1 give very good fit for the average log PDR. But for the log PDR volatility, Model 0 offers the best fit 0.0574,

which is only about 1/7 of the target value 0.40.

The first-order autocorrelation of the log PDR on equity. Compared with the target value (0.90), Model 2 gives the best fit (0.800), and the Model 0 value (0.627) is next. As long-run risks are very persistent and rare events are not, this result is intuitively clear.

Variations in the Value of CRRA γ

Table 1.10 also shows how the asset pricing statistics change with variations in the value of CRRA γ , for $\gamma = 6.3, 7$, and 10. For the RE+LRR model, as γ increases, the mean and volatility of the risk-free rate, and the mean and first- and second-order autocorrelation of the log PDR on equity decrease; the equity premium, the mean and volatility of the market return, and the Sharpe ratio increase. Except for the equity premium and Sharpe ratio, which rise quickly as γ increases, all of the other statistics are relatively insensitive to changes in γ . As a summary, from $\gamma = 6.3$ to 10, the RE+LRR model fits the data less well in general.

The situation of the RE model is similar to that of the RE+LRR model, except for the volatility and the first- and second-order autocorrelation of the log RDR on equity. Similar to the RE+LRR model, from $\gamma = 6.3$ to 10, the RE model fits the data less well in general.

For the LRR model, the statistics about volatilities and the log PDR on equity are relatively insensitive to changes in γ . Unlike the RE+LRR and RE models, from $\gamma = 6.3$ to 10, the LRR model fits the data better in general. The directions of changes in asset pricing statistics match those presented in BY.

Overall, the RE+LRR model renders the smallest value of γ to match the target level of equity premium, and gives the highest estimate for the Sharpe ratio. The predicted

volatilities of the log PDR on equity given by all the three models are significantly smaller than the target value. In general, the three models generate volatilities smaller than the data. Different from the setting in BY, the consumption and dividend claim are assumed to be the same in this study. If an additional dividend claim were introduced to the model and were assumed to be more volatile than the consumption claim as in BY and BKY, the predicted volatilities would be larger.

Variation in the Value of IES ψ

Table 1.11 demonstrates how the asset pricing statistics change when the IES ψ takes values 2, 1.5, and 1.1 for the RE+LRR model: In this process, the predicted mean and volatility of the equity premium, the Sharpe ratio, and the volatility of the log PDR on equity get smaller; the volatility of the risk-free rate and the first- and second-order autocorrelation of the log PDR on equity increase.

Note that the Equation (1.15) is not well defined for IES $\psi = 1$. When $\psi \in (0, 1)$, the iteration of the recursive formula for the PDR of the consumption (output) claim—which is discussed in Section 1.4.1—will go unbounded.

Model Comparison

As I mention in Section 1.3.5, the RE+LRR model is the appropriate model, and I do not do model selection based on how well they fit the asset pricing statistics. But it is informative to compare the asset pricing implications of the three models. Table 1.12 displays the asset pricing statistics calculated by using parameters estimated from the RE+LRR, RE, and LRR models, respectively, together with $\psi = 2$, $\gamma = 6.3$, and $\beta = \exp(-0.031)$. The values

Table 1.11: Asset Pricing Statistics: Data and Model Values for Various IES ψ^\dagger

Moment	Data ³	Model 0 ²	Model 0d ³	Model 0e ⁴
$E(R_f)$	1.009	1.009	1.016	1.024
$E(R_m)$	1.057	1.057	1.055	1.051
$E(R_m - R_f)$	0.048	0.048	0.038	0.027
$\sigma(R_f)$	0.0333	0.0267	0.0345	0.0466
$\sigma(R_m)$	0.1225	0.0683	0.0646	0.0660
$\sigma(R_m - R_f)$	0.1158	0.0670	0.0587	0.0517
$\frac{E(R_m - R_f)}{\sigma(R_m - R_f)}$	0.41	0.719	0.656	0.522
$E(p - d)$	3.30	3.26	3.31	3.40
$\sigma(p - d)$	0.40	0.0574	0.0387	0.0108
$AC1(p - d)$	0.90	0.627	0.630	0.636
$AC2(p - d)$		0.459	0.463	0.470

[†]See Notes of Table 1.10 for explanation on notations.

¹Source: See Table 1.9.

²The same as Column “Model 0” of Table 1.10.

³Calculated using the same parameter values as in Model 0, except for setting $\psi = 1.5$.

⁴Calculated using the same parameter values as in Model 0, except for setting $\psi = 1.1$.

in Column “Model 1c” and “Model 2c” are generally close to the corresponding values in Column “Model 1” and “Model 2” of Table 1.10, respectively, except for the equity premium and the Sharpe ratio in Column “Model 2c.”

Table 1.12: Asset Pricing Statistics: Data and Predicts from the Three Models[†]

Moment	Data ³ ¹	Model 0 ²	Model 1c ³	Model 2c ⁴
$E(R_f)$	1.009	1.009	1.019	1.031
$E(R_m)$	1.057	1.057	1.054	1.049
$E(R_m - R_f)$	0.048	0.048	0.035	0.019
$\sigma(R_f)$	0.0333	0.0267	0.0248	0.0346
$\sigma(R_m)$	0.1225	0.0683	0.0630	0.0608
$\sigma(R_m - R_f)$	0.1158	0.0670	0.0618	0.0633
$\frac{E(R_m - R_f)}{\sigma(R_m - R_f)}$	0.41	0.719	0.559	0.296
$E(p - d)$	3.30	3.26	3.42	3.61
$\sigma(p - d)$	0.40	0.0574	0.0414	0.0479
$AC1(p - d)$	0.90	0.627	0.433	0.735
$AC2(p - d)$		0.459	0.214	0.631

[†]See Notes of Table 1.10 for explanation on notations.

¹Source: See Table 1.9.

²The same as Column "Model 0" of Table 1.10.

³Calculated using parameters estimated from the RE model (see Table 1.7) together with $\psi = 2$, $\gamma = 6.3$, and $\beta = \exp(-0.031)$.

⁴Calculated using parameters estimated from the LRR model (see Table 1.8) together with $\psi = 2$, $\gamma = 6.3$, and $\beta = \exp(-0.031)$.

For the RE model, we see that the equity premium, the Sharpe ratio, and various volatilities in Column “Model 1c” are smaller than the corresponding values in Column “Model 1” of Table 1.10. The main reason for this change is that the RE model gives smaller estimates for the negative short-run and permanent event effects than does the RE+LRR model, as documented in Section 1.3.5.

For the LRR model, compared with the values in Column “Model 2” of Table 1.10, the estimates for the equity premium and Sharpe ratio get larger and closer to the data. The volatilities of the risk-free rate and the log PDR on equity become larger and closer to the data, while those of the market return and the equity premium become smaller and further from the data.

1.5 Conclusion

For the purpose of identifying rare events and long-run risks and evaluating their contributions to asset pricing statistics, this paper proposes an empirical model of rare events and long-run risks, and studies its implications in asset pricing. The major advancements of this study are threefold. First, this study simultaneously identifies the two types of risks in a unified framework. The empirical results on the decomposition of growth gaps reveal the three fundamental distinctions between rare events and long-run risks as stated previously.

Second, this study evaluates the contributions of the two channels to asset pricing statistics, with an emphasis on the equity premium. When the long-run-risk channel is shut down, the estimated model implies that the corresponding equity premium is 4.2%; when the rare-event channel is shut down, the corresponding equity premium is 1.0%. This indicates that a major portion of the equity premium is the compensation for rare events, and

the contribution of long-run risks to equity premium is not significant. Another noticeable result is about the Sharpe ratio. I also compare the asset pricing statistics calculated by using different CRRA γ , IES ψ , and parameter values that are estimated from the RE+LRR, RE, and LRR models, respectively.

Third, this study improves the estimation of disaster process. The estimated *ex ante* annual probability for a country to enter a C (GDP) disaster is 0.0331 (0.0336). For C (GDP) disasters, the average peak-to-trough durations are estimated at 3.2 (3.3) years. The estimated average peak-to-trough drop in C (GDP) is 23%, and on average, about 40% of this decline is reversed in subsequent recoveries. These findings are close to the facts observed in the data used in this study, which demonstrates that the proposed model matches up with the data very well.

Chapter 2

On the Size Distribution of Macroeconomic Disasters¹¹

The coefficient of relative risk aversion, γ , is a key parameter for analyses of behavior toward risk, but good estimates of this parameter do not exist. A promising area for reliable estimation is rare macroeconomic disasters, which have a major influence on the equity premium; see Rietz (1988), Barro (2006), and Barro and Ursúa (2008). For 17 countries with long-term data on returns on stocks and short-term government bills, the average annual (arithmetic) real rates of return were 0.081 on stocks and 0.008 on bills (Barro and Ursúa (2008, Table 5)). Thus, if we approximate the risk-free rate by the average real bill return, the average equity premium was 0.073. An adjustment for leverage in corporate financial structure, using a debt-equity ratio of 0.5, implies that the unlevered equity premium averaged around 0.05.

¹¹Joint with Robert J. Barro, Harvard University. This chapter is based on a paper by the authors published under the same title in *Econometrica*, Vol. 79, No. 5 (Barro and Jin, 2011a).

Previous research (Barro and Ursúa, 2008) sought to explain an equity premium of 0.05 in a representative-agent model calibrated to fit the long-term history of macroeconomic disasters for up to 36 countries. One element in the calibration was the disaster probability, p , measured by the frequency of macroeconomic contractions of magnitude 10% or more. Another feature was the size distribution of disasters, gauged by the observed histogram in the range of 10% and above. Given p and the size distribution, a coefficient of relative risk aversion, γ , around 3.5 accorded with the target equity premium.

The present paper shows that the size distribution of macroeconomic disasters can be characterized by a power law in which the upper-tail exponent, α , is the key parameter. This parametric approach generates new estimates of the coefficient of relative risk aversion, γ , needed to match the target equity premium. We argue that the parametric procedure can generate more accurate estimates than the sample-average approach because of selection problems related to missing data for the largest disasters. In addition, confidence sets for the power-law parameters translate into confidence intervals for the estimates of γ .

Section 2.1 reviews the determination of the equity premium in a representative-agent model with rare disasters. Section 2.2 specifies a familiar, single power law to describe the size distribution of disasters and applies the results to estimate the coefficient of relative risk aversion, γ . Section 2.3 generalizes to a double power law to get a better fit to the observed size distribution of disasters. Section 2.4 shows that the results are robust to reasonable variations in the estimated disaster probability, the target equity premium, and the threshold for disasters (set initially at 10%). Section 2.5 considers possible paradoxes involving an infinite equity premium. Section 2.6 summarizes the principal findings, with emphasis on the estimates of γ .

2.1 The Equity Premium in a Model with Rare Disasters

Barro (2009) worked out the equity premium in a Lucas (1978) tree model with rare but large macroeconomic disasters. (Results for the equity premium are similar in a model with a linear, AK, technology, in which saving and investment are endogenous.) In the Lucas-tree setting, (per capita) real gross domestic product (GDP), Y_t , and consumption, $C_t = Y_t$, evolve as

$$\log(Y_{t+1}) = \log(Y_t) + g + u_{t+1} + v_{t+1}. \quad (2.1)$$

The parameter $g \leq 0$ is a constant that reflects exogenous productivity growth. The random term u_{t+1} , which is independent and identically distributed (i.i.d.) normal with mean 0 and variance σ^2 , reflects “normal” economic fluctuations. The random term v_{t+1} picks up low-probability disasters, as in Rietz (1988) and Barro (2006). In these rare events, output and consumption jump down sharply. The probability of a disaster is the constant $p \geq 0$ per unit of time. In a disaster, output contracts by the fraction b , where $0 < b \leq 1$. The distribution of v_{t+1} is given by

$$\begin{aligned} &\text{probability } 1 - p : v_{t+1} = 0, \\ &\text{probability } p : v_{t+1} = \log(1 - b). \end{aligned}$$

The disaster size, b , follows some probability density function. In previous research, the density for b was gauged by the observed histogram. The present analysis specifies the form of this distribution—as a power law—and estimates the parameters, including the exponent of the upper tail. Note that the expected growth rate, g^* , of consumption and GDP is

$$g^* = g + (1/2) \cdot \sigma^2 - p \cdot Eb,$$

where Eb is the mean disaster size.

Barro (2009) showed that, with a representative agent with Epstein-Zin (Epstein and Zin, 1989) and Weil (1990) preferences, the formula for the unlevered equity premium, when the period length approaches zero, is

$$r^e - r^f = \gamma\sigma^2 + p \cdot E\{b \cdot [(1 - b)^{-\gamma} - 1]\}, \quad (2.2)$$

where r^e is the expected rate of return on unlevered equity (a claim on aggregate consumption flows), r^f is the risk-free rate, and γ is the coefficient of relative risk aversion.¹² The term in curly brackets has a straightforward interpretation under power utility, where γ equals the reciprocal of the intertemporal elasticity of substitution (IES) for consumption. Then this term is the product of the proportionate decline in equity value during a disaster, b , and the excess of marginal utility of consumption in a disaster state compared to that in a normal state, $(1 - b)^{-\gamma} - 1$. Note that in the present setting, the proportionate fall in equity value during a disaster, b , equals the proportionate fall in consumption and GDP during the disaster.

Equation (2.2) can be expressed as

$$r^e - r^f = \gamma\sigma^2 + p \cdot [E(1 - b)^{-\gamma} - E(1 - b)^{1-\gamma} - Eb]. \quad (2.3)$$

Equation (2.3) shows that the key properties of the distribution of b are the expectations of the variable $1/(1 - b)$ taken to the powers γ and $\gamma - 1$. (The Eb term has a minor impact.)

Barro and Ursúa (2008) studied macroeconomic disasters by using long-term annual

¹²The present analysis assumes that the representative agent's relative risk aversion is constant. Empirical support for this familiar specification appears in Brunnermeier and Nagel (2008) and Chiappori and Paiella (2011).

data for real per capita consumer expenditure, C , for 24 countries and real per capita GDP (henceforth, called GDP) for 36 countries.¹³ These data go back at least to 1914 and as far back as 1870, depending on availability, and end in 2006. The annual time series, including sources, are available at <http://rbarro.com/data-sets/>.

Barro and Ursúa (2008) followed Barro (2006) by using an NBER (National Bureau of Economic Research) -style peak-to-trough measurement of the sizes of macroeconomic contractions. Starting from the annual time series, proportionate contractions in C and GDP were computed from peak to trough over periods of 1 or more years, and declines by 10% or greater were considered. This method yielded 99 disasters for C (for 24 countries) and 157 for GDP (36 countries). The average disaster sizes, subject to the threshold of 10%, were similar for the two measures: 0.215 for C and 0.204 for GDP. The mean durations of the disasters were also similar: 3.6 years for C and 3.5 years for GDP. The list of the disaster events—by country, timing, and size—is given in Barro and Ursúa (2008, Tables C1 and C2).¹⁴

Equation (2.1) is best viewed as applying to short periods, approximating continuous time. In this setting, disasters arise as downward jumps at an instant of time, and the disaster size, b , has no time units. In contrast, the underlying data on C and GDP are annual flows. In relating the data to the theory, there is no reason to identify disaster sizes, b , with large contractions in C or GDP observed particularly from one year to the next. In

¹³This approach assumes that the same process for generating macroeconomic disasters and the same model of household risk aversion apply to all countries at all points in time. In general, reliable estimation of parameters for a rare-disasters model requires a lot of data coming from a population that can be viewed as reasonably homogeneous. However, Barro and Ursúa (2008) found that results on the determinants of the equity premium were similar if the sample were limited to Organization for Economic Cooperation and Development (OECD) countries.

¹⁴These data on disaster sizes are the ones used in the current study, except for a few minor corrections. The values used are in the Supplemental Material (Barro and Jin, 2011b).

fact, the major disaster events—exemplified by the world wars and the Great Depression—clearly feature cumulative declines over several years, with durations of varying length. In Barro (2006) and Barro and Ursúa (2008), the disaster jump sizes, b , in the continuous-time model—corresponding to equation (2.3) for the equity premium—were approximated empirically by the peak-to-trough measures of cumulative, proportionate decline. Barro (2006, Section V) showed that this procedure would be reasonably accurate if the true model were one with discrete periods with length corresponding to the duration of disasters (all with the same duration, say $3\frac{1}{2}$ years).

The peak-to-trough method for gauging disaster sizes has a number of shortcomings, addressed in the research by NSBU. In this work, the underlying model features a probability per year, p , of entering into a disaster state. (Disaster events are allowed to be correlated across countries, as in the world wars and the Great Depression.) Disasters arise in varying sizes (including occasional bonanzas), and the disaster state persists stochastically. This specification generates frequency distributions for the cumulative size and duration of disasters. In addition, as in Gourio (2008), post-disaster periods can feature recoveries in the form of temporarily high growth rates.¹⁵

The most important implication of NSBU for the equity premium comes from the recoveries. Since, on average, only half the decline in consumption during a disaster turns out to be permanent, the model's predicted equity premium falls short of the value in equation (2.3). The other extensions have less influence on the equity premium, although

¹⁵One nonissue (raised by Constantinides (2008, pp. 343–344) and Donaldson and Mehra (2008, p. 84) is the apparent mismatch between the units for rates of return—per year—and the measurement of disaster sizes by cumulative declines over multiple years (with a mean duration around $3\frac{1}{2}$ years). As already noted, the peak-to-trough measures of macroeconomic decline are approximations to the model's jump declines, which have no time units.

the stochastic duration of disasters matters because of effects on the correlation between consumption growth and stock returns during disasters.

The present analysis uses the peak-to-trough measures of declines in C and GDP to generate an empirical distribution of disaster sizes, b . Figures 2.1 and 2.2 show the corresponding histograms for transformed disaster sizes, $1/(1-b)$, for C and GDP , respectively. The findings in NSBU suggest that these measures will be satisfactory for characterizing the distribution of disaster sizes, but that some downward adjustment to the equity premium in equation (2.3) would be appropriate to account particularly for the partly temporary nature of the typical disaster.

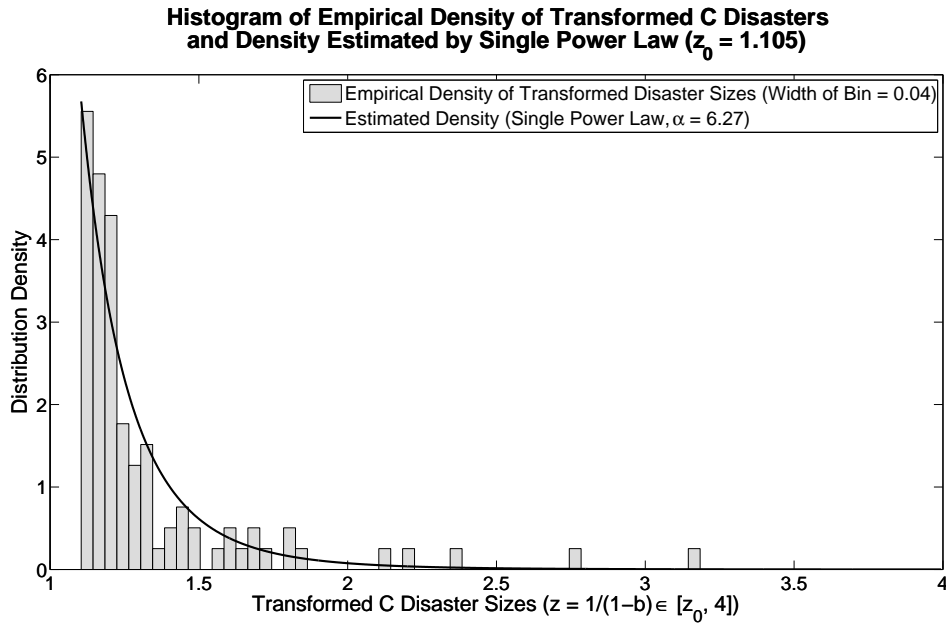


Figure 2.1: Histogram of the empirical density of transformed C disasters and the density estimated by the single power law

Notes: The threshold is $z_0 = 1.105$, corresponding to $b = 0.095$. For the histogram, multiplication of the height on the vertical axis by the bin width of 0.04 gives the fraction of the total observations (99) that fall into the indicated bin. The results for the single power law for C , shown by the curve, correspond to Table 2.1.

As in previous research, the estimated disaster probability, p , equals the ratio of the

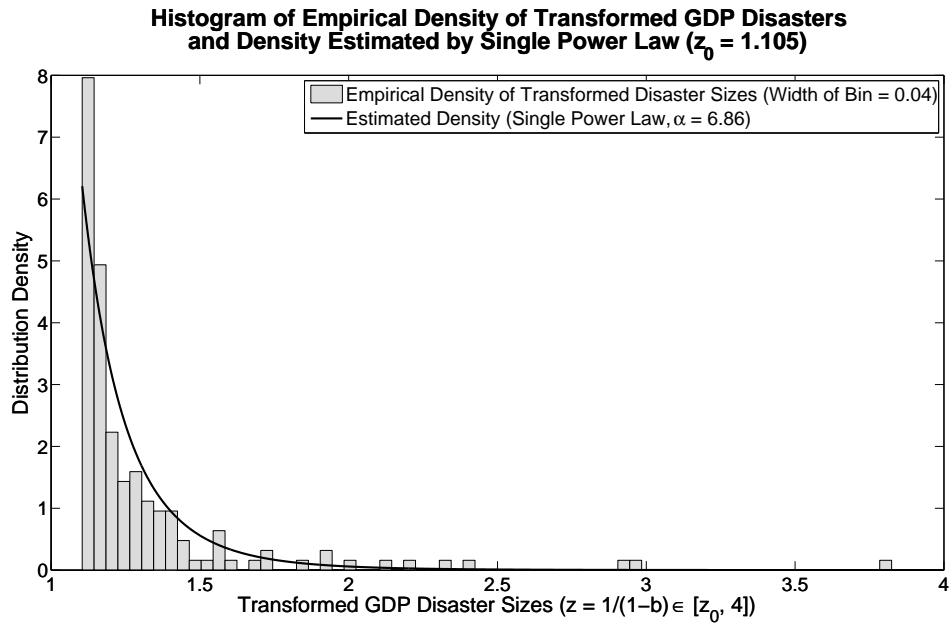


Figure 2.2: Histogram of the empirical density of transformed GDP disasters and the density estimated by the single power law

Notes: The threshold is $z_0 = 1.105$, corresponding to $b = 0.095$. For the histogram, multiplication of the height on the vertical axis by the bin width of 0.04 gives the fraction of the total observations (157) that fall into the indicated bin. The results for the single power law for GDP, shown by the curve, correspond to Table 2.1.

number of disasters to the number of nondisaster years. This calculation yields $p = 0.0380$ per year for C and $p = 0.0383$ for GDP. Thus, disasters (macroeconomic contractions of 10% or more) typically occur around three times per century. The United States experience for C is comparatively mild, featuring only two contractions of 10% or more over 137 years—with troughs in 1921 and 1933. However, for GDP, the U.S. data show five contractions of 10% or more, with troughs in 1908, 1914, 1921, 1933, and 1947.¹⁶

Barro and Ursúa (2008, Tables 10 and 11) used the observed histograms for disaster sizes from the C and GDP data (Figures 2.1 and 2.2) to compute the expectation (that is,

¹⁶The 1947 GDP contraction was associated with the demobilization after World War II and did not involve a decline in C. The 1908 and 1914 GDP contractions featured declines in C, but not up to the threshold of 10%.

the sample average) of the expression in brackets on the right side of equation (2.3) for alternative coefficients of relative risk aversion, γ . The resulting values were multiplied by the estimated p to calculate the disaster term on the right side of the equation. The other term on the right side, $\gamma\sigma^2$, was computed under the assumption $\sigma = 0.02$ per year. However, as in Mehra and Prescott (1985), this term was trivial, compared to the equity premium of around 0.05, for plausible values of γ (even with higher, but still reasonable, values of σ). Hence, the disaster term ended up doing almost all the work in explaining the equity premium. A key finding was that a γ around 3.5 got the model's equity premium into the neighborhood of the target value of 0.05.

2.2 Single-Power-Law Distribution

We work with the transformed disaster size

$$z \equiv 1/(1 - b),$$

which is the ratio of normal to disaster consumption or GDP. This variable enters into the formula for the equity premium in equation (2.3). The threshold for b , taken to be 0.095, translates into one for z of $z_0 = 1.105$. As b approaches 1, z approaches ∞ , a limiting property that accords with the usual setting for a power-law distribution.

We start with a familiar, single power law, which specifies the density function as

$$f(z) = Az^{-(\alpha+1)} \tag{2.4}$$

for $z \geq z_0$, where $A > 0$ and $\alpha > 0$. The condition that the density integrate to 1 from z_0 to

∞ implies

$$A = \alpha z_0^\alpha. \quad (2.5)$$

The power-law distribution in equation (2.4) has been applied widely in physics, economics, computer science, ecology, biology, astronomy, and so on. For a review, see Mitzenmacher (2004a). Gabaix (2009) provided examples of power laws in economics and finance, and discussed forces that can generate these laws. The examples include sizes of cities (Gabaix and Ioannides, 2004), stock-market activity (Gabaix, Gopikrishnan, Plerou and Stanley, 2003, 2006), chief executive officer compensation (Gabaix and Landier, 2008), and firm size (Luttmer, 2007). The power-law distribution has been given many names, including heavy-tail distribution, Pareto distribution, Zipfian distribution, and fractal distribution.

Pareto (1897) observed that, for large populations, a graph of the logarithm of the number of incomes above a level x against the logarithm of x yielded points close to a straight line with slope $-\alpha$. This property corresponds to a density proportional to $x^{-(\alpha+1)}$; hence, Pareto's α corresponds to ours in equation (2.4). The straight-line property in a log-log graph can be used to estimate α , as was done by Gabaix and Ibragimov (2011) using least squares. A more common method uses maximum-likelihood estimation (MLE), following Hill (1975). We use MLE in our study.

In some applications, such as the distribution of income, the power law gives a poor fit to the observed frequency data over the whole range, but provides a good fit to the upper tail.¹⁷ In many of these cases, a double power law—with two different exponents

¹⁷There have been many attempts to explain this Paretian tail behavior, including Champernowne (1953), Mandelbrot (1960), and Reed (2003).

over two ranges of z —fits the data well. For uses of this method, see Reed (2003) on the distribution of income and Mitzenmacher (2004b) on computer file sizes. The double power law requires estimation of a cutoff value, δ , for z , above which the upper-tail exponent, α , for the usual power law applies. For expository purposes, we begin with the single power law, but problems in fitting aspects of the data eventually motivate a switch to the richer specification.

The single-power-law density in equations (2.4) and (2.5) implies that the equity premium in equation (2.3) is given by

$$r^e - r^f = \gamma\sigma^2 + p \cdot \left\{ \left(\frac{\alpha}{\alpha - \gamma} \right) z_0^\gamma - \left(\frac{\alpha}{\alpha + 1 - \gamma} \right) z_0^{\gamma-1} + \left(\frac{\alpha}{\alpha + 1} \right) \cdot \left(\frac{1}{z_0} \right) - 1 \right\} \quad (2.6)$$

if $\alpha > \gamma$. (This formula makes no adjustment for the partially temporary nature of disasters, as described earlier.) For given p and z_0 , the disaster term on the right side involves a race between γ , the coefficient of relative risk aversion, and α , the tail exponent. An increase in γ raises the disaster term, but a rise in α implies a thinner tail and, therefore, a smaller disaster term. If $\alpha \leq \gamma$, the tail is sufficiently thick that the equity premium is infinite. This result corresponds to a risk-free rate, r^f , of $-\infty$. We discuss these possibilities later. For now, we assume $\alpha > \gamma$.

We turn now to estimation of the tail exponent, α . When equation (2.4) applies, the log likelihood for N independent observations on z (all at least as large as the threshold, z_0) is

$$\log(L) = N \cdot [\alpha \cdot \log(z_0) + \log(\alpha)] - (\alpha + 1) \cdot [\log(z_1) + \cdots + \log(z_N)], \quad (2.7)$$

where we used the expression for A from equation (2.5). The MLE condition for α follows

readily as

$$N/\alpha = \log(z_1/z_0) + \cdots + \log(z_N/z_0). \quad (2.8)$$

We obtained standard errors and 95% confidence intervals for the estimate of α from bootstrap methods.¹⁸

Table 2.1 shows that the point estimate of α for the 99 C disasters is 6.27, with a standard error of 0.81 and a 95% confidence interval of (4.96, 8.12). Results for the 157 GDP disasters are similar: the point estimate of α is 6.86, with a standard error of 0.76 and a 95% confidence interval of (5.56, 8.48).

Given an estimate for α —and given $\sigma = 0.02$, $z_0 = 1.105$, and a value for p (0.0380 for C and 0.0383 for GDP)—we need only a value for γ in equation (2.6) to determine the predicted equity premium, $r^e - r^f$. To put it another way, we can find the value of γ needed to generate $r^e - r^f = 0.05$ for each value of α . (The resulting γ has to satisfy $\gamma < \alpha$ for $r^e - r^f$ to be finite.) In Table 2.1, the point estimate for α of 6.27 from the single power law for the C data requires $\gamma = 3.97$. The corresponding standard error for the estimated γ is 0.51, with a 95% confidence interval of (3.13, 5.13). For the GDP data, the point estimate of γ is 4.33, with a standard error of 0.48 and a 95% confidence interval of (3.50, 5.33).

To assess these results, we now evaluate the fit of the single power law. Figure 2.1 compares the histogram for the C disasters with the frequency distribution implied by the single power law in equations (2.4) and (2.5), using $z_0 = 1.105$ and $\alpha = 6.27$ from Table 2.1. An important inference is that the single power law substantially underestimates the frequency of large disasters. Similar results apply for GDP in Figure 2.2.

¹⁸See Efron and Tibshirani (1993). We get similar results based on $-2 \cdot \log(\text{likelihood ratio})$ being distributed asymptotically as a chi-squared distribution with 1 degree of freedom (Greene, 2012, see).

Table 2.1: Single and Double Power Laws; Threshold is $z_0 = 1.105^\dagger$

Parameter	Point Estimate	Standard Error	95% Confidence Interval
C Data (99 disasters)			
Single power law			
α	6.27	0.81	(4.96, 8.12)
γ	3.97	0.51	(3.13, 5.13)
Double power law			
α	4.16	0.87	(2.66, 6.14)
β	10.10	2.40	(7.37, 15.17)
δ	1.38	0.13	(1.24, 1.77)
γ	3.00	0.52	(2.16, 4.15)
GDP Data (157 disasters)			
Single power law			
α	6.86	0.76	(5.56, 8.48)
γ	4.33	0.48	(3.50, 5.33)
Double power law			
α	3.53	0.97	(2, 39, 6.07)
β	10.51	3.81	(8.67, 20.98)
δ	1.47	0.15	(1.21, 1.69)
γ	2.75	0.56	(2.04, 4.21)

[†]The single power law, given by equations (2.4) and (2.5), applies to transformed disaster sizes, $z \equiv 1/(1 - b)$, where b is the proportionate decline in C (real personal consumer expenditure per capita) or real GDP (per capita). Disasters are at least as large as the threshold, $z_0 = 1.105$, corresponding to $b \geq 0.095$. The table shows the maximum-likelihood estimate of the tail exponent, α . The standard error and 95% confidence interval come from bootstrap methods. The corresponding estimates of γ , the coefficient of relative risk aversion, come from calculating the values needed to generate an unlevered equity premium of 0.05 in equation (2.6) (assuming $\sigma = 0.02$ and $p = 0.0380$ for C and 0.0383 for GDP). For the double power law, given by equations (2.10)–(2.12), the table shows the maximum-likelihood estimates of the two exponents, α (above the cutoff) and β (below the cutoff), and the cutoff value, δ . The corresponding estimates of γ come from calculating the values needed to generate an unlevered equity premium of 0.05 in a more complicated version of equation (2.6).

The failures in the single power law are clearer in diagrams for cumulative densities. The straight lines in Figures 2.3 and 2.4 show, for C and GDP, respectively, fitted logs of probabilities that transformed disaster sizes exceed the values shown on the horizontal axes. The lines connecting the points show logs of normalized ranks of disaster sizes (as in Gabaix and Ibragimov (2011)). If the specified single power law were valid, the two graphs in each figure should be close to each other over the full range of z . However, the figures demonstrate that the single power laws underestimate the probabilities of being far out in the upper tails.

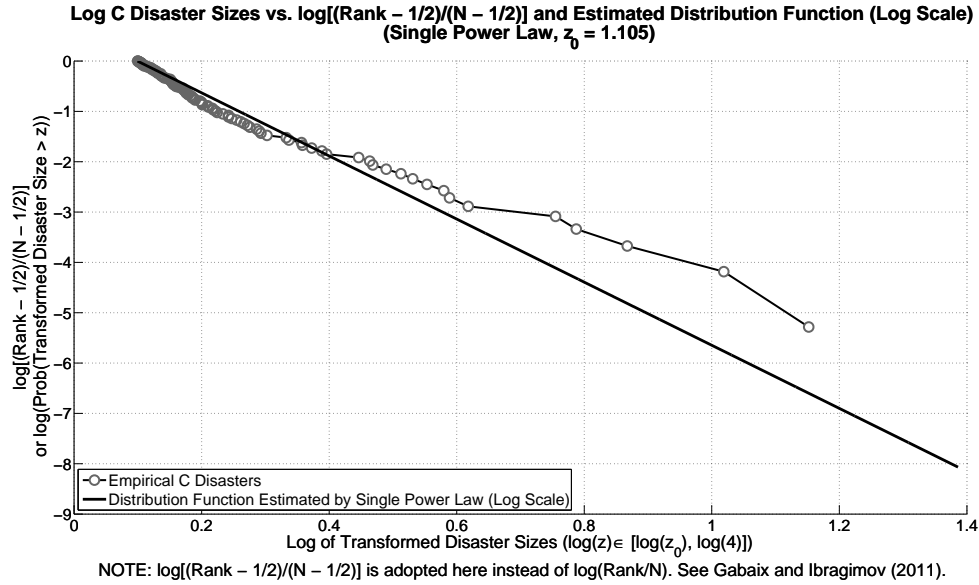


Figure 2.3: Estimated log-scale tail distribution and log of transformed ranks of C disaster sizes versus log of transformed C disaster sizes

Notes: The straight line corresponding to the log-scale tail distribution comes from the estimated single power law for C in Table 2.1. The ranks of the disaster sizes are transformed as $\log[(\text{rank} - 1/2)/(N - 1/2)]$, in accordance with Gabaix and Ibragimov (2011). The lines connecting these points should—if the estimated power law is valid—converge pointwise in probability to the log-scale tail distribution, as N approaches infinity.

One way to improve the fits is to allow for a smaller tail exponent at high disaster sizes by generalizing to a double power law. This form specifies an upper-tail exponent,

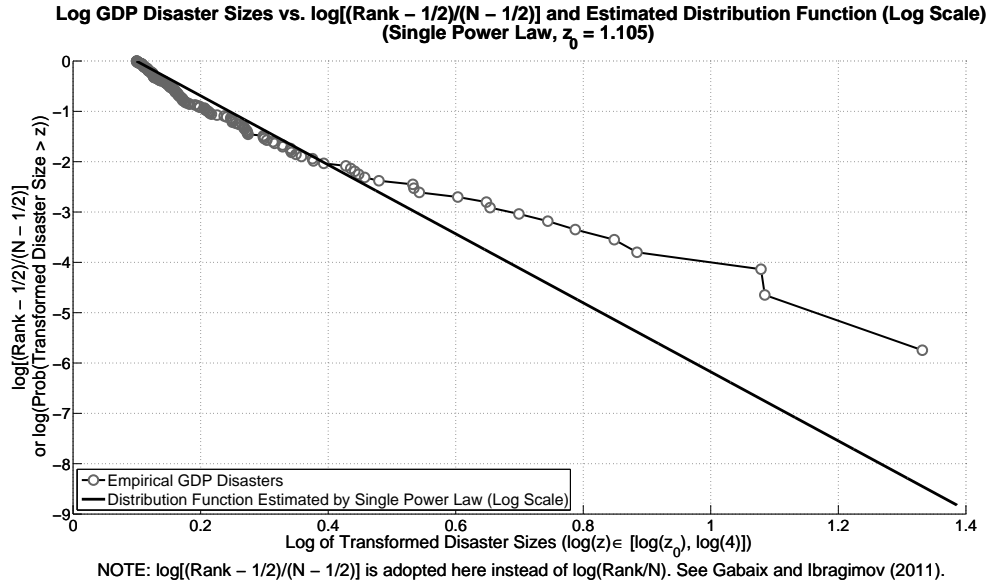


Figure 2.4: Estimated log-scale tail distribution and log of transformed ranks of GDP disaster sizes versus log of transformed GDP disaster sizes

Notes: The straight line corresponding to the log-scale tail distribution comes from the estimated single power law for GDP in Table 2.1. See note to Figure 2.3 for further information.

α , that applies for z at or above a cutoff value, $\delta \geq z_0$, and a lower-tail exponent, β , that applies below the cutoff value for $z_0 \leq z < \delta$. This generalization requires estimation of three parameters: the exponents, α and β , and the cutoff, δ . We still treat the threshold, z_0 , as known and equal to 1.105. We should note that the critical parameter for the equity premium is the upper-tail exponent, α . The lower-tail exponent, β , is unimportant; in fact, the distribution need not follow a power law in the lower part. However, we have to specify a reasonable form for the lower portion to estimate the cutoff, δ , which influences the estimate of α .

2.3 Double-Power-Law Distribution

The double-power-law distribution, with exponents β and α , takes the form

$$f(z) = \begin{cases} 0, & \text{if } z < z_0, \\ Bz^{-(\beta+1)}, & \text{if } z_0 \leq z < \delta, \\ Az^{-(\alpha+1)}, & \text{if } \delta \leq z, \end{cases} \quad (2.9)$$

where $\beta, \alpha > 0$, $A, B > 0$, $z_0 > 0$ is the known threshold, and $\delta \geq z_0$ is the cutoff separating the lower and upper parts of the distribution. The conditions that the density integrate to 1 over $[z_0, \infty)$ and that the densities be equal just to the left and right of δ imply

$$B = A\delta^{\beta-\alpha}, \quad (2.10)$$

$$\frac{1}{A} = \frac{\delta^{\beta-\alpha}}{\beta} (z_0^{-\beta} - \delta^{-\beta}) + \frac{\delta^{-\alpha}}{\alpha}. \quad (2.11)$$

The single power law in equations (2.4) and (2.5) is the special case of equations (2.9)–(2.11) when $\beta = \alpha$.

The position of the cutoff, δ , determines the number, K , among the total observations, N , that lie below the cutoff. The remaining $N - K$ observations are at or above the cutoff. Therefore, the log likelihood can be expressed as a generalization of equation (2.7) as

$$\begin{aligned} \log(L) = & N \cdot \log(A) + K \cdot (\beta - \alpha) \cdot \log(\delta) \\ & - (\beta + 1) \cdot [\log(z_1) + \cdots + \log(z_K)] \\ & - (\alpha + 1) \cdot [\log(z_{K+1}) + \cdots + \log(z_N)], \end{aligned} \quad (2.12)$$

where A satisfies equation (2.11).

We use maximum likelihood to estimate α , β , and δ . One complication is that small changes in δ cause discrete changes in K when one or more observations lie at the cutoff.

These jumps do not translate into jumps in $\log(L)$ because the density is equal just to the left and right of the cutoff. However, jumps arise in the derivatives of $\log(L)$ with respect to the parameters. This issue does not cause problems in finding numerically the values of (α, β, δ) that maximize $\log(L)$ in equation (2.12). Moreover, we get virtually the same answers if we rely on the first-order conditions for maximizing $\log(L)$ calculated while ignoring the jump problem for the cutoff. These first-order conditions are generalizations of equation (2.8).¹⁹

In Table 2.1, the sections labeled double power law show the point estimates of (α, β, δ) for the C and GDP data. We again compute standard errors and 95% confidence intervals using bootstrap methods. A key finding is that the upper-tail exponent, α , is estimated to be much smaller than the lower-tail exponent, β . For example, for C, the estimate of α is 4.16, standard error equal to 0.87, with a confidence interval of (2.66, 6.14), whereas that for β is 10.10, standard error equal to 2.40, with a confidence interval of (7.37, 15.17). The estimates reject the hypothesis $\alpha = \beta$ in favor of $\alpha < \beta$ at low p -values (for C and GDP).

Table 2.1 shows that the estimated cutoff value, δ , for the C disasters is 1.38; recall that this value corresponds to the transformed disaster size, $z \equiv 1/(1 - b)$. The corresponding cutoff for b is 0.275. With this cutoff, 77 of the C crises fall below the cutoff, whereas 22 are above. The corresponding cutoff for b with the GDP crises is 0.320, implying that 136 events fall below the cutoff, whereas 21 are above. Despite the comparatively small number

¹⁹The expressions are

$$\begin{aligned} \frac{1}{\alpha} &= \left(\frac{1}{N-K} \right) \cdot \left[\log \left(\frac{z_{K+1}}{\delta} \right) + \cdots + \log \left(\frac{z_N}{\delta} \right) \right], \\ \alpha \cdot \left[\log \left(\frac{z_{K+1}}{z_0} \right) + \cdots + \log \left(\frac{z_N}{z_0} \right) \right] &+ \beta \cdot \left[\log \left(\frac{z_1}{z_0} \right) + \cdots + \log \left(\frac{z_K}{z_0} \right) \right] = N, \\ \frac{\delta}{z_0} &= \left[\frac{N\alpha + K \cdot (\beta - \alpha)}{\alpha \cdot (N-K)} \right]^{\frac{1}{\beta}}. \end{aligned}$$

of crises above the cutoffs, we know from previous research Barro and Ursúa (2008, Tables 10 and 11) that the really large crises have the main influence on the equity premium. That assessment still holds for the present analysis.

Figure 2.5 compares the histogram for the C disasters with the frequency distribution implied by the double power law in equations (2.9)–(2.11), using $z_0 = 1.105$, $\alpha = 4.16$, $\beta = 10.10$, and $\delta = 1.38$ from Table 2.1. Unlike the single power law in Figure 2.1, the double power law accords well with the histogram. Results are similar for the GDP data (not shown). Figures 2.6 and 2.7 provide corresponding information for cumulative densities. Compared with the single power laws in Figures 2.3 and 2.4, the double power laws accord much better with the upper-tail behavior. The improved fits suggest that the double power law would be superior for estimating the coefficient of relative risk aversion, γ .

With respect to the equity premium, the key difference in Table 2.1 between the double and single power laws is the substantially smaller upper-tail exponents, α . Since the estimated α is now close to 4, rather than exceeding 6, the upper tails are much fatter when gauged by the double power laws. These fatter tails mean that a substantially lower coefficient of relative risk aversion, γ , accords with the target equity premium of 0.05.

Equation (2.3) still determines the equity premium, $r^e - r^f$. For given γ , a specification of (α, β, δ) , along with $z_0 = 1.105$, determines the relevant moments of the disaster-size distribution. That is, we get a more complicated version of equation (2.6). (As before, this formulation does not adjust for the partially temporary nature of macroeconomic disasters.) Crucially, a finite $r^e - r^f$ still requires $\alpha > \gamma$. The results determine the estimate of γ that corresponds to those for (α, β, δ) in Table 2.1 (still assuming $\sigma = 0.02$ and $p = 0.0380$ for C and 0.0383 for GDP). This procedure yields point estimates for γ of 3.00 from the C

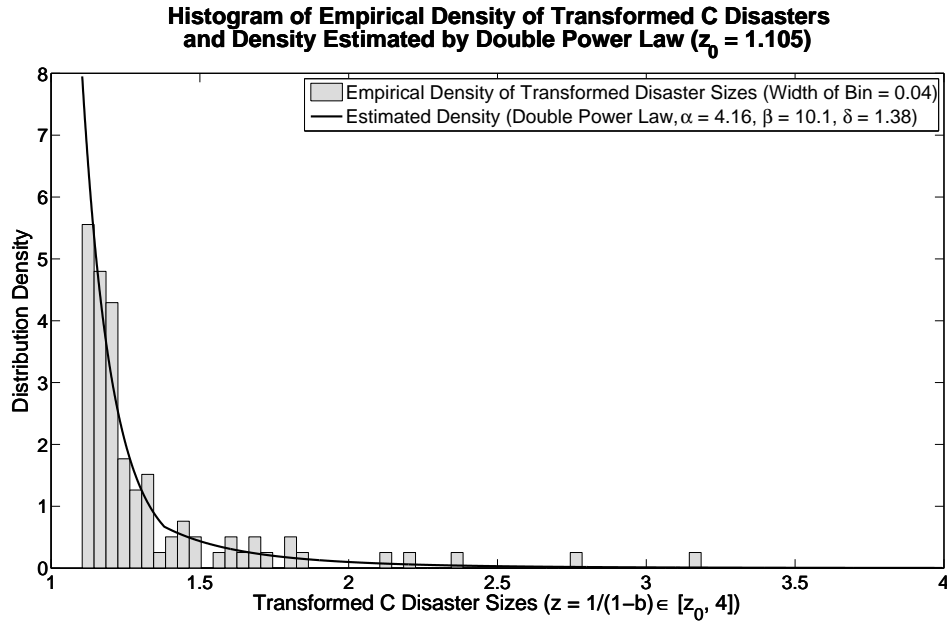
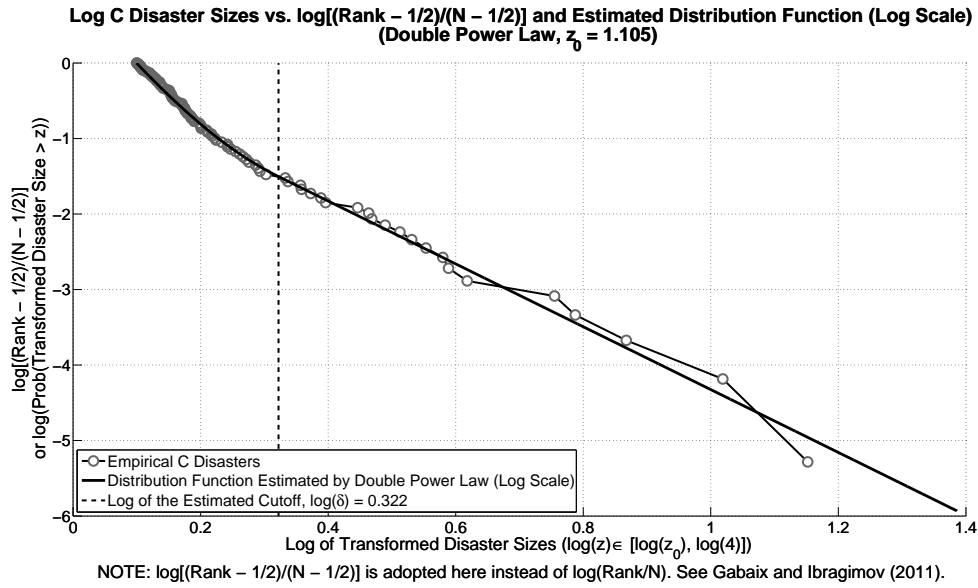


Figure 2.5: Histogram of the empirical density of transformed C disasters and the density estimated by the double power law ($z_0 = 1.105$)

Notes: For the histogram, multiplication of the height shown on the vertical axis by the bin width of 0.04 gives the fraction of the total observations (99) that fall into the indicated bin. The results for the double power law for C, shown by the curve, are based on Table 2.1.

disasters and 2.75 from the GDP disasters.

As before, we use bootstrap methods to determine standard errors and 95% confidence intervals for the estimates of γ . Although the main parameter that matters is the upper-tail exponent, α , we allow also for variations in β and δ . For the C disasters, the estimated γ of 3.00 (Table 2.1) has a standard error of 0.52, with a 95% confidence interval of (2.16, 4.15). For GDP, the estimate of 2.75 has a standard error of 0.56, with a confidence interval of (2.04, 4.21). Thus, γ is estimated to be close to 3, with a 95% confidence band of roughly 2



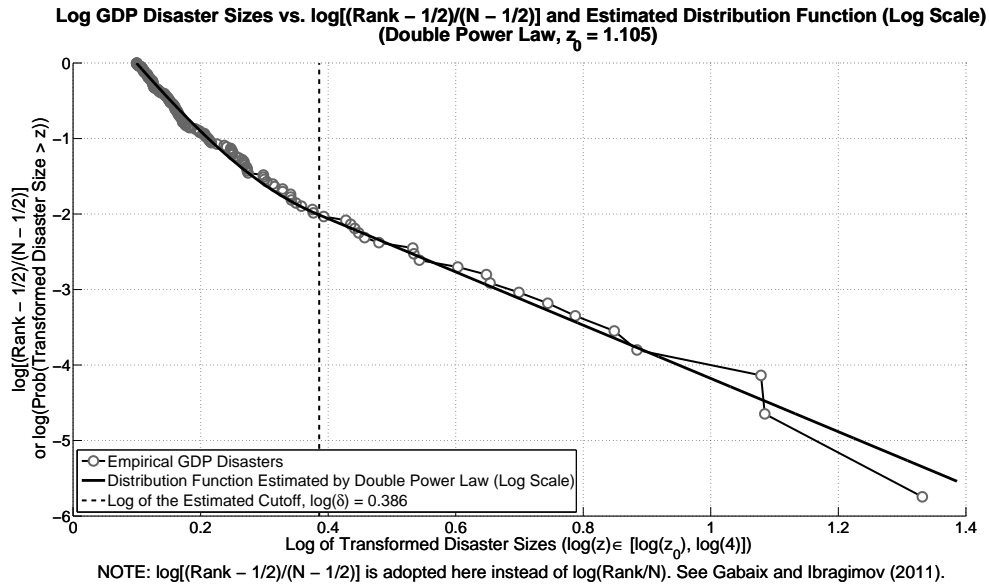


Figure 2.7: Estimated log-scale tail distribution and log of transformed ranks of GDP disaster sizes versus log of transformed GDP disaster sizes

Notes: The line with two segments corresponding to the log-scale tail distribution comes from the estimated double power law for GDP in Table 2.1. See note to Figure 2.3 for further information.

The last comparison reflects interesting differences in the two methods: the moments of the size distribution that determine the equity premium in equation (2.3) can be estimated from a parametric form (such as the double power law) that accords with the observed distribution of disaster sizes or from sample averages of the relevant moments (corresponding to histograms). A disadvantage of the parametric approach is that misspecification of the functional form—particularly for the far upper tails that have few or no observations—may give misleading results. In contrast, sample averages seem to provide consistent estimates for any underlying functional form. However, the sample average approach is sensitive to a selection problem, whereby data tend to be missing for the largest disasters (sometimes because governments have collapsed or are fighting wars). This situation must apply to an end-of-world (or, at least, end-of-country) scenario, discussed later, where $b = 1$. The

tendency for the largest disasters to be missing from the sample means that the sample-average approach tends to underestimate the fatness of the tails, thereby leading to an overstatement of γ .²¹ In contrast, the parametric approach (with a valid functional form) may be affected little by missing data in the upper tail. That is, the estimate of the upper-tail exponent, α , is likely to have only a small upward bias due to missing extreme observations, which have to be few in number. This contrast explains why our estimated γ around 3 from the double power laws (Table 2.1) is noticeably smaller than the value around 3.5 generated by the observed histograms.

2.4 Variations in Disaster Probability, Target Equity Premium, and Threshold

We consider now whether the results on the estimated coefficient of relative risk aversion, γ , are robust to uncertainty about the disaster probability, p , the target equity premium, $r^e - r^f$, and the threshold, z_0 , for disaster sizes. For p , the estimate came from all the sample data, not just the disasters: p equaled the ratio of the number of disasters (for C or GDP) to the number of nondisaster years in the full sample. Thus, a possible approach to assessing uncertainty about the estimate of p would be to use a model that incorporates all the data, along the lines of NSBU. We could also consider a richer setting in which p varies over time, as in Gabaix (2012). We carry out here a more limited analysis that assesses how

²¹The magnitude of this selection problem has diminished with Ursua's (2010) construction of estimates of GDP and consumer spending for periods, such as the world wars, where standard data were missing. Recent additions to his data set—not included in our current analysis—are Russia, Turkey, and China (for GDP). As an example, the new data imply that the cumulative contraction in Russia from 1913 to 1921 was 62% for GDP and 71% for C.

“reasonable” variations in p influence the estimates of γ .²²

Figure 2.8 gives results for C , and analogous results apply for GDP (not shown). Recall that the baseline value for p of 0.038 led to an estimate for γ of 3.00, with a 95% confidence interval of (2.16, 4.15). Figure 2.8 shows that lowering p by a full percentage point (to 0.028) increases the point estimate of γ to 3.2, whereas raising p by a full percentage point (to 0.048) decreases the point estimate of γ to 2.8. Thus, substantial variations in p have only a moderate effect on the estimated γ .

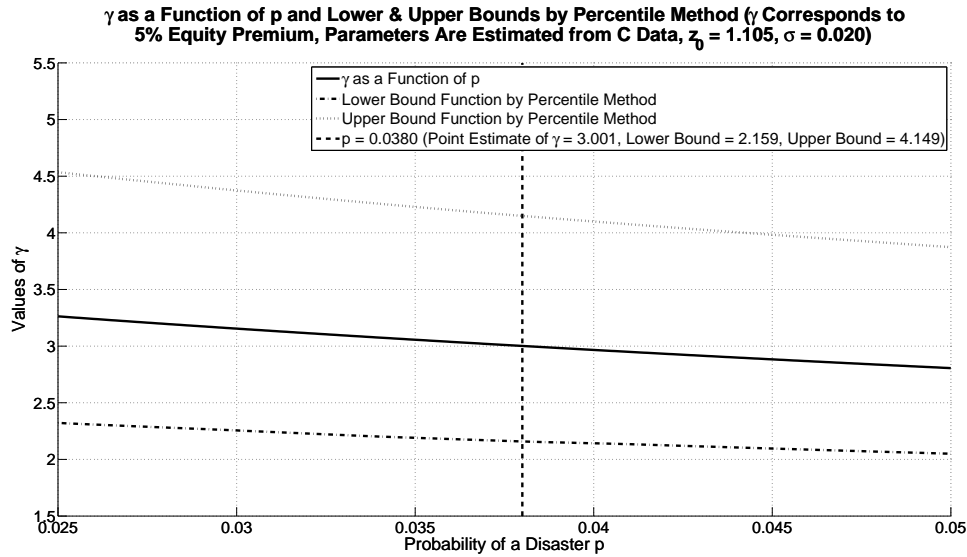


Figure 2.8: Estimates of the coefficient of relative risk aversion, γ , for alternative disaster probabilities, given target equity premium of 0.05, threshold of $z_0 = 1.105$, and $\sigma = 0.020$

Notes: These results correspond to the estimated double power law for C in Table 2.1.

We assumed that the target equity premium was 0.05. More realistically, there is uncertainty about this premium, which can also vary over time and space (due, for example, to

²²For a given set of observed disaster sizes (for C or GDP), differences in p do not affect the maximum-likelihood estimates for the parameters of the power-law distributions. We can think of differences in p as arising from changes in the overall sample size while holding fixed the realizations of the number and sizes of disaster events.

shifts in the disaster probability, p). As with our analysis of p , we consider how reasonable variations in the target premium influence the estimated γ . An allowance for a higher target equity premium is also a way to adjust the model to account for the partly temporary nature of macroeconomic disasters. That is, since equation (2.3) overstates the model's equity premium when the typical disaster is partly temporary (as described before), an increase in the target premium is a way to account for this overstatement.

Equation (2.3) shows that variations in the equity premium, $r^e - r^f$, on the left side are essentially equivalent, but with the opposite sign, to variations in p on the right side. Therefore, diagrams for estimates of γ versus $r^e - r^f$ look similar to Figure 2.8, except that the slope is positive. Quantitatively, for the C data, if $r^e - r^f$ were 0.03 rather than 0.05, the point estimate of γ would be 2.6 rather than 3.0. On the other side, if $r^e - r^f$ were 0.07, the point estimate of γ would be 3.2. Results with GDP are similar. Thus, substantial variations in the target equity premium have only a moderate influence on the estimated γ .

The results obtained thus far apply for a fixed threshold of $z_0 = 1.105$, corresponding to proportionate contractions, b , of size 0.095 or greater. This choice of threshold is arbitrary. In fact, our estimation of the cutoff value, δ , for the double power laws in Table 2.1 amounts to endogenizing the threshold that applies to the upper tail of the distribution. We were able to estimate δ by MLE because we included in the sample a group of observations that potentially lie below the cutoff. Similarly, to estimate the threshold, z_0 , we would have to include observations that potentially lie below the threshold. As with estimates of p , this extension requires consideration of all (or at least more of) the sample, not just the disasters.

As in the analysis of disaster probability and target equity premium, we assess the

impact of variations in the threshold on the estimated coefficient of relative risk aversion, γ . We consider a substantial increase in the threshold, z_0 , to 1.170, corresponding to $b = 0.145$, the value used in Barro (2006). This rise in the threshold implies a corresponding fall in the disaster probability, p (gauged by the ratio of the number of disasters to the number of nondisaster years in the full sample). For the C data, the number of disasters declines from 99 to 62, and p decreases from 0.0380 to 0.0225. For the GDP data, the number of disasters falls from 157 to 91 and p declines from 0.0383 to 0.0209. That is, the probability of a disaster of size 0.145 or more is about 2% per year, corresponding to roughly two events per century.

The results in Table 2.2, for which the threshold is $z_0 = 1.170$, can be compared with those in Table 2.1, where $z_0 = 1.105$. For the single power law, the rise in the threshold causes the estimated exponent, α , to adjust toward the value estimated before for the upper part of the double power law (Table 2.1). Since the upper-tail exponents (α) were lower than the lower-tail exponents (β), the estimated α for the single power law falls when the threshold rises. For the C data, the estimated α decreases from 6.3 in Table 2.1 to 5.5 in Table 2.2, and the confidence interval shifts downward accordingly. The reduction in α implies that the estimated γ declines from 4.0 in Table 2.1 to 3.7 in Table 2.2, and the confidence interval shifts downward correspondingly. Results for the single power law for GDP are analogous.²³

²³These results apply even though the higher threshold reduces the disaster probability, p . That is, disaster sizes in the range between 0.095 and 0.145 no longer count. As in Barro and Ursúa (2008, Tables 10 and 11), the elimination of these comparatively small disasters has only a minor impact on the model's equity premium and, hence, on the value of γ required to generate the target premium of 0.05. The more important force is the thickening of the upper tail implied by the reduction of the tail exponent, α .

Table 2.2: Single and Double Power Laws with Higher Threshold, $z_0 = 1.170^{\dagger}$

Parameter	Point Estimate	Standard Error	95% Confidence Interval
C Data (62 disasters)			
Single power law			
α	5.53	0.85	(4.16, 7.61)
γ	3.71	0.53	(2.83, 4.97)
Double power law			
α	4.05	0.87	(2.81, 6.12)
β	11.36	8.27	(6.63, 39.78)
δ	1.37	0.15	(1.21, 1.86)
γ	3.00	0.54	(2.21, 4.29)
GDP Data (91 disasters)			
Single power law			
α	5.67	0.81	(4.39, 7.49)
γ	3.86	0.51	(3.03, 4.99)
Double power law			
α	4.77	1.00	(2.42, 6.24)
β	59.22	22.13	(7.90, 76.73)
δ	1.20	0.17	(1.20, 1.75)
γ	3.41	0.60	(2.04, 4.34)

[†]See the notes to Table 2.1. Disasters are now all at least as large as the threshold $z_0 = 1.170$, corresponding to $b \geq 0.145$. The disaster probability, p , is now 0.0225 for C and 0.0209 for GDP.

With a double power law, the change in the threshold has much less impact on the estimated upper-tail exponent, α , which is the key parameter for the estimated γ . For the C data, the rise in the threshold moves the estimated α from 4.16 in Table 2.1 to 4.05 in Table 2.2, and the confidence interval changes correspondingly little.²⁴ These results imply that the results for γ also change little, going from a point estimate of 3.00 with a confidence interval of (2.16, 4.15) in Table 2.1 to 3.00 with an interval of (2.21, 4.29) in Table 2.2. Results for GDP are analogous. We conclude that a substantial increase in the threshold has little effect on the estimated γ .

2.5 Can the Equity Premium Be Infinite?

Weitzman (2007), building on Geweke (2001), argued that the equity premium can be infinite (and the risk-free rate minus infinite) when the underlying shocks are log normally distributed with unknown variance. In this context, the frequency distribution for asset pricing is the t -distribution, for which the tails can be sufficiently fat to generate an infinite equity premium. The potential for an infinite equity premium arises also—perhaps more transparently—in our setting based on power laws.

For a single power law, the equity premium, $r^e - r^f$, in equation (2.6) rises with the coefficient of relative risk aversion, γ , and falls with the tail exponent, α , because a higher α implies a thinner tail. A finite equity premium requires $\alpha > \gamma$, and this condition still applies with a double power law, with α representing the upper-tail exponent. Thus, it is easy to generate an infinite equity premium in the power-law setting. For a given γ , the

²⁴The rise in the threshold widens the confidence interval for the estimated lower-tail exponent, β . As the threshold rises toward the previously estimated cutoff, δ , the lower tail of the distribution becomes increasingly less relevant.

tail has only to be sufficiently fat; that is, α has to satisfy $\alpha \leq \gamma$.

However, we assume that the equity premium, $r^e - r^f$, equals a known (finite) value, 0.05. The important assumption here is not that the premium equals a particular number, but rather that it lies in an interval of something like 0.03–0.07 and is surely not infinite. Our estimation, therefore, assigns no weight to combinations of parameters, particularly of α and γ , that generate a counterfactual premium, such as ∞ . For given α (and the other parameters), we pick (i.e., estimate) γ to be such that the premium equals the target, 0.05. Estimates constructed this way always satisfy $\alpha > \gamma$ and, therefore, imply a finite equity premium.

The successful implementation of this procedure depends on having sufficient data so that there are enough realizations of disasters to pin down the upper-tail exponent, α , within a reasonably narrow range. Thus, it is important that the underlying data set is very large in a macroeconomic perspective: 2963 annual observations on consumer expenditure, C , and 4653 on GDP. Consequently, the numbers of disaster realizations—99 for C and 157 for GDP—are sufficient to generate reasonably tight confidence intervals for the estimates of α .

Although our underlying data set is much larger than those usually used to study macroeconomic disasters, even our data cannot rule out the existence of extremely low probability events of astronomical size. Our estimated disaster probabilities, p , were 3.8% per year for C and GDP, and the estimated upper-tail exponents, α , were close to 4 (Table 2.1). Suppose that there were a far smaller probability p^* , where $0 < p^* \ll p$, of experiencing a super disaster; that is, one drawn from a size distribution with a much fatter tail, characterized by an exponent α^* , where $0 < \alpha^* \ll \alpha$. If p^* is extremely low, say 0.01% per

year, there is a good chance of seeing no realizations of super disasters even with 5000 observations. Thus, our data cannot rule out the potential for these events, and these far-out possibilities may matter. In particular, regardless of how low p^* is, to fit the target equity premium of 0.05, the coefficient of relative risk aversion, γ , has to satisfy $\gamma < \alpha^*$ to get a finite equity premium.²⁵ If α^* can be arbitrarily low (a possibility not ruled out by direct observation when p^* is extremely low), the estimated γ can be arbitrarily close to zero. We, thus, get a reversal of the Mehra-Prescott (1985) puzzle, where the coefficient of relative risk aversion required to match the observed equity premium was excessive by a couple orders of magnitude.²⁶

Any upper bound $B < 1$ on the potential disaster size, b , would eliminate the possibility of an infinite equity premium. In this sense, the extreme results depend on the possibility of an end-of-the-world event, where $b = 1$. To consider this outcome, suppose now that the very small probability p^* refers only to $b = 1$. In this case, it is immediate from equation (2.3) that the equity premium, $r^e - r^f$, is infinite if $\gamma > 0$. Thus, with the assumed form of utility function,²⁷ any positive probability of apocalypse (which cannot be ruled out by “data”), when combined with an equity premium around 0.05, is inconsistent with a positive degree of risk aversion.

The reference to an end-of-the-world event suggests a possible resolution of the puzzle.

²⁵The assumption here, perhaps unreasonable, is that constant relative risk aversion applies arbitrarily far out into the tail of low consumption.

²⁶This reversal is the counterpart of the one described in Weitzman (2007, p. 1110): “Should we be trying to explain the *puzzle pattern*: why is the actually observed equity premium so embarrassingly *high* while the actually observed riskfree rate is so embarrassingly *low*...? Or should we be trying to explain the opposite *antipuzzle pattern*: why is the actually observed equity premium so embarrassingly *low* while the actually observed riskfree rate is so embarrassingly *high*...?”

²⁷The result does not depend on the constant relative risk aversion form, but only on the condition that the marginal utility of consumption approaches infinity as consumption tends toward zero.

The formula for the equity premium in equation (2.2) involves a comparison of the return on equity, interpreted as a claim on aggregate consumption, with that on a risk-free asset, interpreted as a short-term government bill. However, no claim can deliver risk-free consumption (from whom and to whom?) once the world has ended. Therefore, at least in the limit, we have to allow for risk in the “risk-free” claim.

Even if we restrict to $b < 1$, a disaster that destroys a large fraction, b , of consumption is likely to generate partial default on normally low-risk assets such as government bills. Empirically, this low return typically does not involve explicit default but rather high inflation and, thereby, low realized real returns on nominally denominated claims during wartime (see Barro (2006, Section I.c)). For the 99 C crises considered in the present analysis, we have data (mainly from Global Financial Data) on real bill returns for 58, of which 33 were during peacetime and 25 involved wars. The median realized real rates of return on bills (arithmetic) were 0.014 in the peacetime crises, similar to that for the full sample, and -0.062 in the wartime crises. Thus, the main evidence for partial default on bills comes from wars that involved macroeconomic depressions.

To generalize the model (without specifically considering war versus peace), suppose that the loss rate on government bills is $\Phi(b)$, where $0 \leq \Phi(b) \leq 1$. We assume $\Phi(0) = 0$, so that bills are risk-free in normal times. The formula for the equity premium in equation (2.2) becomes

$$r^e - r^f = \gamma\sigma^2 + p \cdot E\{[b - \Phi(b)] \cdot [(1 - b)^{-\gamma} - 1]\}. \quad (2.13)$$

Thus, instead of the loss rate, b , on equity, the formula involves the difference in the loss rates during disasters on equity versus bills, $b - \Phi(b)$. We previously assumed $\Phi(b) = 0$, but a more reasonable specification is $\Phi'(b) \geq 0$, with $\Phi(b)$ approaching 1 as b approaches

1. The equity premium in equation (2.13) will be finite if, as b approaches 1, $b - \Phi(b)$ approaches 0 faster than $(1 - b)^{-\gamma}$ approaches infinity. In particular, the marginal utility of consumption (for a hypothetical survivor) may be infinite if the world ends ($b = 1$), but the contribution of this possibility to the equity premium can be nil because no asset can deliver consumption once the world has disappeared.

2.6 Summary of Main Findings

The coefficient of relative risk aversion, γ , is a key parameter for analyses of behavior toward risk. We estimated γ by combining information on the probability and sizes of macroeconomic disasters with the observed long-term average equity premium. Specifically, we calculated what γ had to be to accord with a target unlevered equity premium of 5% per year within a representative-agent model that allows for rare disasters.

In our main calibration, based on the long-term global history of macroeconomic disasters, the probability, p , of disaster (defined as a contraction in per capita consumption or GDP by at least 10% over a short period) is 3.8% per year. The size distribution of disasters accords well with a double power law, with an upper-tail exponent, α , of about 4. The resulting estimate of γ is close to 3, with a 95% confidence interval of 2 to 4. This finding is robust to whether we consider consumer expenditure or GDP and to variations in the estimated disaster probability, p , the target equity premium, and the threshold for the size distribution. The results can also accommodate seemingly paradoxical situations in which the equity premium may appear to be infinite.

Chapter 3

Estimation and Testing of ARMA Models with Root Cancellation

3.1 Introduction

The (near) root cancellation situation in autoregressive moving average (ARMA) models is an important example of weak identification/nonidentification. The simplest special case, the ARMA(1,1) model with (near) root cancellation, is of both theoretical and empirical significance, and is studied by researchers from various aspects, for instance, Ansley and Newbold (1980), Andrews and Ploberger (1996), Kleibergen and Hoek (2000), Nelson and Startz (2007), Ma and Nelson (2008), and Cogley and Startz (2012). Andrews and Cheng (2012a) provide a unified treatment of a class of models where lack of identification and weak identification occurs in part of the parameter space, and apply their results to examples including the ARMA(1,1) case.

This paper studies the asymptotic behavior of the maximum likelihood estimator (MLE)

for ARMA(1, 1) models with (near) root cancellation in the full range of strength-of-identification scenarios. To investigate the finite-sample behavior of the estimator under weak identification, we consider its asymptotic behavior under *drifting sequences of parameters (true distributions)*, as in Andrews and Cheng (2012a), Staiger and Stock (1997), Stock and Wright (2000), and many other researches on weak instruments. In this study, we give a complete classification of asymptotic identification categories for all the drifting sequences of parameters. Compared with previous works, the classification here is more complete and reasonable, and much more detailed information about the asymptotic distributions of MLE is provided in this study, including the analytical characterization of asymptotic distributions for the full range of strengths of identification and the detailed formula for the semi-strong identification case. This study also reveals how the strength of identification of parameters change with T , the sample size, and δ , the sum of the AR and MA parameters.

Another contribution of this paper is to propose a novel statistic, called J -statistic, for conducting joint tests on the AR and MA parameters. The J -statistic is straightforward to calculate and has a standard normal limiting distribution, which is asymptotically pivotal, and the corresponding J -test is robust to the full range of identification strengths (i.e., non, weak, semi-strong, and strong identification). The simulation shows that the actual test size is very close to the nominal value (5%), and the test power is satisfactory in the δ direction. The performance of J -test is particularly good when the true parameter values are of or close to the root cancellation cases, which is a desirable property.

Consider the following sequence of *causal invertible* ARMA(1,1) processes. For each $T \geq 1$, the T -th ARMA(p, q) process takes the form

$$(1 - \phi_T L)Y_{t,T} = (1 + \theta_T L)\sigma_T \epsilon_{t,T}, \quad (3.1)$$

for $t \in \mathbb{Z}$, where $\sigma_T > 0$, L is the *lag operator*, $\{Y_{t,T}\}_{t=1-p}^T$ are observed r.v.'s (random variables) and $\{\epsilon_{t,T}\}_t$ are unobserved innovations and are (independent and identically distributed) i.i.d.(0,1) with finite fourth moment.

Given the true parameter $(\zeta^{*'}, \sigma^*)'$ of the model, where $\zeta^* = (\phi^* + \theta^*, \theta^*)' \triangleq (\delta^*, \theta^*)'$, the issue of *identification/nonidentification* may occur at the population level (either in the strict sense or in the sense of identification by a criterion function). In this case, it seems that the issue of identification/nonidentification is *unrelated* to sample size. However, given a drifting sequence $\{\zeta_T\}_{T=1}^\infty = \{(\phi_T + \theta_T, \theta_T)'\}_{T=1}^\infty \triangleq \{(\delta_T, \theta_T)'\}_{T=1}^\infty$, it turns out that its asymptotic identification/nonidentification depends on the rate of convergence of $\{\delta_T\}_{T=1}^\infty$. Thus, the asymptotic identification/nonidentification is closely related to sample size T . Based on the classification results in this study, it is convenient to define three types of drifting sequences of parameters as follows.

Definition 3.1.1. A drifting sequence $\{\zeta_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ (defined in (3.22)) is called asymptotically root-cancelling or simply root-cancelling, if $\lim_{T \rightarrow \infty} \delta_T = 0$. A root-cancelling drifting sequence $\{\zeta_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ is called

1. a $1/2-\langle 0 \rangle$ sequence, if $\lim_{T \rightarrow \infty} T^{1/2} \delta_T = 0$;
2. a $1/2-\langle 1 \rangle$ sequence, if it is convergent with $\lim_{T \rightarrow \infty} T^{1/2} \delta_T = c \in \mathbb{R} \setminus \{0\}$;
3. a $1/2-\langle \infty \rangle$ sequence, if it is convergent with $\lim_{T \rightarrow \infty} T^{1/2} \delta_T = \pm\infty$.

It is clear that the three types of sequences are *mutually exclusive*, and Theorem 3.1.3 shows that if a drifting sequence $\{\zeta_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ with $\lim_{T \rightarrow \infty} \delta_T = 0$ is not of one of the above three types, then the asymptotic limit of the distribution of ML estimator $\hat{\zeta}_T$

doesn't exist. Also note that, for $1/2-\langle 0 \rangle$ sequences, the definition focuses on the rate of convergence of $\{\delta_T\}_T$, and there is no requirement for the convergence of $\{\xi_T\}_{T=1}^\infty$.

Let $\Delta\hat{\xi}_T = \hat{\xi}_T - \xi_T$, and the convergence rate of $\Delta\hat{\xi}_T$ is a key issue of this study. Generally, given a sequence $\{x_T\}_{T=1}^\infty$ of r.v.'s (including nonrandom quantities), the ROC (rate of convergence) of $\{x_T\}_{T=1}^\infty$ (to 0) is described by a nonrandom nonnegative sequence $\{a_T\}_{T=1}^\infty$, i.e., we use notations like $x_T = O_p(a_T)$ and $x_T = o_p(a_T)$.²⁸

For the case of ARMA(1,1), given a sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ and the optimization parameter space $\Xi_{1,1}^+$ (defined in (3.21)), if asymptotic distributions of $\hat{\delta}_T$, $\hat{\theta}_T$, and $\hat{\xi}_T$ exist, we denote them by $\pi_{\hat{\delta}_T}(\{\xi_T\}_{T=1}^\infty | \Xi_{1,1}^+)$, $\pi_{\hat{\theta}_T}(\{\xi_T\}_{T=1}^\infty | \Xi_{1,1}^+)$, and $\pi_{\hat{\xi}_T}(\{\xi_T\}_{T=1}^\infty | \Xi_{1,1}^+)$, respectively. When there is no ambiguity, we also use notations such as $\pi_{\hat{\delta}_T}(\{\xi_T\}_{T=1}^\infty)$ and $\pi_{\hat{\theta}_T}$, etc.

Table 3.1 summaries the results about *asymptotic identification categories* for θ_T for root-cancellation situation in ARMA(1,1) models, which presents a complete description of strength of identification for a full range of drifting sequences of true distributions. It turns out that the strength of identification of θ_T is determined by the race between the ROC of $\{\delta_T\}_T$ and $T^{-1/2}$. Andrews and Cheng (2012a) provide a table (Table 1) somewhat similar to Table 3.1 here. Roughly speaking, the classification in Andrews and Cheng (2012a) is coarser than the one presented here. One significant difference between the two classifications is that, according to results in this paper, any $1/2-\langle 0 \rangle$ sequence belongs to the unidentified category; while in Table 1 of Andrews and Cheng (2012a), it falls into the weakly identified category. Also note that much more sequences are listed in the unidentified category in the table here. In fact, the results in this paper are much more inclusive

²⁸ It is important to note that we allow entries of $\{a_T\}_{T=1}^\infty$ to be 0. If $a_{T'} = 0$ for some $T' \geq 1$, $x_T = O_p(a_T)$ and $x_T = o_p(a_T)$ simply mean that the corresponding $x_{T'} = 0$ (a.e.). See also Footnote 37 and 45 for the use of the "big-O" (O) and "little-o" (o) notations in this paper. Footnote 30 is related to the treatment here.

as they provide a complete classification of drifting sequences $\{\tilde{\zeta}_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ in terms of asymptotic distributions. Thus, the classification presented here is more reasonable and complete. Moreover, there is much more detailed information on ROC's and asymptotic distributions for various types of drifting sequences of parameters, including theorems about the existence of the asymptotic distribution and the equivalence classes of drifting sequences.

For many results in this paper, it is assumed that the drifting sequence of parameters converges, as in most of the related literature. For example, the condition

$$\lim_{T \rightarrow \infty} \tilde{\zeta}_T = \tilde{\zeta}_0 = (\delta_0, \theta_0)' \quad (3.2)$$

is often assumed.²⁹ However, some results in the paper don't require the convergence of the corresponding drifting sequence at all, e.g., Theorem 3.2.11. Also, the $1/2\text{-}\langle 0 \rangle$ sequences need not to be convergent. Thus, in this paper, a drifting sequence of parameters may not converge unless it is said so explicitly.

Three main theorems in this study are listed below, where the terms like C_T and G_T , etc., are defined in later sections.

Theorem 3.1.2 (Asymptotic Distributions). *Given a drifting sequence $\{\tilde{\zeta}_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, we have:*

1. *For $1/2\text{-}\langle 0 \rangle$ sequences, the MLE $\hat{\theta}_T$ is not consistent and θ_T is asymptotically unidentified.*

For all these sequences, the asymptotic distribution $\pi_{\hat{\zeta}_T}$ is the same, and is determined by

²⁹When the Gaussian (quasi) likelihood function is adopted, parameter σ_T is actually irrelevant in the ML estimation of $\tilde{\zeta}_T$ (see Section 3.2 for an explanation). The results in this paper won't change no matter whether $\lim_{T \rightarrow \infty} \sigma_T$ exists. For this reason, parameter σ_T is often omitted in the discussion.

Table 3.1: Asymptotic Identification Categories for θ_T in ARMA(1,1) Models

CATEGORY	$\{\beta_T\}$ SEQUENCE	NOTES [†]
Unidentified	1/2- $\langle 0 \rangle$ sequences	$\pi_{\hat{\xi}_T}$ exists for any of these sequences, even if $\lim \theta_T$ doesn't exist. Furthermore, $\pi_{\hat{\xi}_T}$ is the same for all these sequences (Case (1) of Theorem 3.1.2), and $\Delta \hat{\theta}_T = O_p(1)$.
Weakly Identified	1/2- $\langle 1 \rangle$ sequences	$\pi_{\hat{\xi}_T}$ is given in Case (2) of Theorem 3.1.2, and $\Delta \hat{\theta}_T = O_p(1)$.
Semi-strongly Identified	1/2- $\langle \infty \rangle$ sequences	$\pi_{\hat{\xi}_T}$ is bivariate normal and is given in Case (3) of Theorem 3.1.2, and $\Delta \hat{\theta}_T = O_p((T^{1/2}\delta_T)^{-1})$.
Strongly Identified	sequences with $\lim \tilde{\xi}_T = \xi_0 = (\delta_0, \theta_0)'$, $\delta_0 \neq 0$	The usual strongly identified case, and $\Delta \hat{\xi}_T = O_p(T^{-1/2})$ (Theorem 3.1.3).
No Asymptotic Distribution	other sequences	Theorem 3.1.3.

[†]For any sequence $\{\tilde{\xi}_T\}_{T=1}^\infty$ whose asymptotic distribution exists, $\Delta \hat{\delta}_T = O_p(T^{-1/2})$ (Corollary 3.2.20).

shape analysis function

$$\mathcal{S}_T(\theta|0, \phi_T) = 2(1 - \theta^2)C_T(\theta)G_T(\theta) \quad (3.3)$$

and (3.73).

2. For a $1/2-\langle 1 \rangle$ sequence with $\lim_{T \rightarrow \infty} d_T = d_0 \in \mathbb{R} \setminus \{0\}$, the MLE $\hat{\theta}_T$ is not consistent and θ_T is weakly identified. The asymptotic distribution $\pi_{\hat{\xi}_T}$ is determined by shape analysis function

$$\mathcal{S}_T(\theta|d_0, \phi_0) = 2 \left[(1 - \theta^2)C_T(\theta) + \frac{(1 - \theta^2)d_0}{1 + \phi_0\theta} \right] [G_T(\theta) + d_0h(\theta|\phi_0)] \quad (3.4)$$

and (3.73).

3. For a $1/2-\langle \infty \rangle$ sequence, the MLE $\hat{\theta}_T$ is consistent and θ_T is semi-strongly unidentified. In this case, on a sequence of sets whose measure approaching 1, the solution to the first-order equation is unique and the FOC becomes the necessary and sufficient condition for minima. The asymptotic distribution $\pi_{\hat{\xi}_T}$ is determined asymptotically by Equation $\mathcal{R}_T(\theta|d_T, \phi_T) = 0$ and (3.73), where

$$\mathcal{R}_T(\theta|d_T, \phi_T) \triangleq G_T(\theta) + \frac{d_T}{(1 + \phi_T\theta_T)(1 - \theta_T^2)^2}(\theta - \theta_T). \quad (3.5)$$

More specifically, we have

$$\begin{pmatrix} T^{1/2} \Delta \hat{\delta}_T \\ d_T \Delta \hat{\theta}_T \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \mathcal{V}(\theta_0)), \text{ as } T \rightarrow \infty, \quad (3.6)$$

where $\mathcal{V}(\theta_0) = (1 - \theta_0^2) \begin{pmatrix} (1 + \theta_0^2) & \theta_0(1 - \theta_0^2) \\ \theta_0(1 - \theta_0^2) & (1 - \theta_0^2)^2 \end{pmatrix}$.

Theorem 3.1.3 (Existence of the Asymptotic Distribution). *Given a drifting sequence $\{\tilde{\xi}_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, the asymptotic distribution of $\hat{\xi}_T$ exists if and only if it is a $1/2-\langle i \rangle$ sequence ($i = 0, 1, \infty$) or a convergent sequence with $\lim_{T \rightarrow \infty} \tilde{\xi}^* = (\delta^*, \theta^*)'$ and $\delta^* \neq 0$. The asymptotic distribution of $\hat{\xi}_T$ for the latter is given in Corollary B.1.5.*

Theorem 3.1.4 (Equivalence Classes of Drifting Sequences). *Let \mathcal{D} denote the set of drifting sequences in $\Xi_{1,1}^*$. We define a binary relation “ \mid ” between the drifting sequences. For two sequences $\{\xi_T\}_{T=1}^\infty = \{(\delta_T, \theta_T)'\}_{T=1}^\infty, \{\xi_T^\diamond\}_{T=1}^\infty = \{(\delta_T^\diamond, \theta_T^\diamond)'\}_{T=1}^\infty \in \mathcal{D}$, we say $\{\xi_T\}_{T=1}^\infty \mid \{\xi_T^\diamond\}_{T=1}^\infty$ if one of the following conditions are satisfied*

1. *Both sequences do not have asymptotic distributions,*
2. *Both are $1/2$ - $\langle 0 \rangle$ sequences,*
3. *$\lim_{T \rightarrow \infty} \xi_T = \lim_{T \rightarrow \infty} \xi_T^\diamond = (\delta^*, \theta^*)'$, where $\delta^* \neq 0$,*
4. *$\lim_{T \rightarrow \infty} \xi_T = \lim_{T \rightarrow \infty} \xi_T^\diamond = (0, \theta^*)'$, and $\{\delta_T\}_T$ and $\{\delta_T^\diamond\}_T$ are equivalent infinitesimals.³⁰*

Then “ \mid ” is an equivalence relation. Moreover, if $\{\xi_T\}_{T=1}^\infty \mid \{\xi_T^\diamond\}_{T=1}^\infty$, then either they have the same asymptotic distribution, or both of them do not have an asymptotic distribution.

We call the method presented in this paper the *global approach* as it explores the asymptotic behavior of the estimator by checking the “global picture” of the likelihood function on the entire parameter space, instead of analyzing it locally with some expansion around some point like those commonly used to studying asymptotics. In general, suppose we estimate parameter θ by minimizing a criterion function $\tilde{Q}_T(\theta)$ over a parameter space Θ , where T is the sample size. The key idea of global approach is to analyze the behavior of the random function $\tilde{Q}_T(\theta)$ over Θ “globally.” In other words, the global analysis does not appeal to local expansions of one kind or another, or the local expansions do not play a central role. Instead, the results about uniform convergence are central in this approach. At the first glimpse, this approach may seem to be complicated. However, it turns out that

³⁰Note that we allow the entries of $\{\delta_T\}_T$ and $\{\delta_T^\diamond\}_T$ to be 0 for generality. See Footnote 28 for details.

this approach has distinct advantages for handling the complication caused by the highly nonlinear feature of the root cancellation situation in the ARMA models.

First, the global analysis applies to all identification categories, and is a natural choice to explore the strength of identification for a full range of drifting sequences. As we do not rely on the power of local expansions, there is no essential difference between no, weak, semi-strong and strong identification from the viewpoint of global approach, although different identification strengths may require slightly different techniques to handle them. The case of ARMA model indicates that the global approach involves calculation of more terms as the identification strength intensifies. Moreover, the global approach intuitively reveals the mechanism that determines the strengths of identification, and provides theoretical results on the series of “continuously” changing asymptotic distributions. Thus, the approach is also “global” in the sense of examining the full spectrum of strengths of identification.

Second, the power of the global approach is in some sense “flexible.” When studying the behavior of random function $\tilde{Q}_T(\theta)$ over Θ , whether or not to calculate certain terms depends on the goal of the research. In theory, every term can be calculated using this approach, so that the calculation can be completely exact. In the case of ARMA(1,1), it turns out that we do not have to compute every term for studying the no, weak, and semi-strong identification, and the use of O_p - and o_p - terms greatly alleviate the burden of the calculation. In other words, the no, weak, and semi-strong identification cases are even easier to deal with using the global approach than the strong identification case, which is contrary to the situation of using other approaches. If we also want to study the usual strongly identified case, then more terms need to be calculated. We choose not to do that

in this study, as the usual strongly identified case has been well studied elsewhere.

The rest of the paper is organized as follows. Section 3.2 shows the detailed derivation of the asymptotic behavior of the MLE in the ARMA(1,1) model with root cancellation. Section 3.3 presents the J -statistic, J -test, and the simulation results.

For notational simplicity, we sometimes write a function $f(c)$ with $c = (a', b')'$ as $f(a, b)$ instead of $f(c)$, where a and b are vectors. In this paper, \mathbb{N} stands for the set of natural numbers starting from 0, and \mathbb{Z}_+ stands for the set of positive integers starting from 1.

3.2 MLE in the ARMA(1,1) Model

3.2.1 Preliminaries

The Equation (3.1) is equivalent to

$$Y_{t,T} = (1 - \phi_T L)^{-1} (1 + \theta_T L) \sigma_T \epsilon_{t,T}, \quad (3.7)$$

and

$$(1 + \theta_T L)^{-1} (1 - \phi_T L) Y_{t,T} = \sigma_T \epsilon_{t,T}. \quad (3.8)$$

For a typical ARMA(1,1) process, root cancellation and *degeneracy* are not considered, i.e., we have

$$-\phi_T \neq \theta_T \quad (3.9)$$

and

$$\phi_T \theta_T \neq 0. \quad (3.10)$$

Let

$$\beta_T = (\phi_T, \theta_T)'. \quad (3.11)$$

Based on the above requirements, the parameter space for β_T , the AR and MA parameters of an ARMA(1,1) process, is usually defined as

$$B_{1,1}^\circ \triangleq \{\beta = (\phi, \theta)' \in \mathbb{R}^2 \mid |\phi|, |\theta| \in (0, 1), -\phi \neq \theta\}. \quad (3.12)$$

We may also define

$$B_{1,1} \triangleq \{\beta = (\phi, \theta)' \in \mathbb{R}^2 \mid |\phi_k|, |\theta_{k'}| < 1, -\phi \neq \theta\} \quad (3.13)$$

to allow the degenerate ARMA(1,1) cases, so that the discussion is more general. It is clear that $B_{1,1}^\circ$ is open in \mathbb{R}^2 , and for every $\beta^* \in B_{1,1}^\circ$, the true parameter β^* (of the corresponding ARMA(1,1) process $Y_{\beta^*} = \{Y_{t,\beta^*}\}_t$) is *identifiable* (in the strict sense or in the sense of identification by a criterion function, e.g., the likelihood criterion function here). Actually, the parameter β^* is *strongly identified* (see Proposition B.1.4), i.e., the estimator $\tilde{\beta}_T$ has the following two properties

$$(i) \ T^{1/2}(\tilde{\beta}_T - \beta^*) \xrightarrow{d} \mathcal{N}(0, V), \text{ and } (ii) \text{ there exists } \tilde{V} \xrightarrow{p} V, \quad (3.14)$$

as $T \rightarrow \infty$.

In this study, we focus on the case of *root cancellation* where the AR and MA roots are asymptotically equal. Other situations, like the *unit root* case, are not considered here. Thus, we have

$$\delta_T \triangleq \phi_T + \theta_T \rightarrow 0, \text{ as } T \rightarrow \infty. \quad (3.15)$$

When $\lim_{T \rightarrow \infty} \beta_T$ exists, we write

$$\lim_{T \rightarrow \infty} \beta_T = \beta_0 = (\phi_0, \theta_0)', \quad (3.16)$$

and by (3.15),

$$\phi_0 = -\theta_0. \quad (3.17)$$

It is also assumed that the limits $|\phi_0|, |\theta_0| < 1$, so that the complication incurred by the unit root cases is avoided.

It turns out that it is more convenient to reparametrize $\beta_T = (\phi_T, \theta_T)'$ as

$$\xi_T = (\delta_T, \theta_T)', \quad (3.18)$$

where $\delta_T = \phi_T + \theta_T$. After the reparametrization, if we relax the restriction that $-\phi \neq \theta$, $B_{1,1}$ becomes

$$\tilde{\Xi}_{1,1} \triangleq \{\xi = (\delta, \theta)' | \delta \in (-1 + \theta, 1 + \theta), |\theta| < 1\}. \quad (3.19)$$

For any $\xi \in \tilde{\Xi}_{1,1}$, the corresponding β must be contained in $\overline{B_{1,1}^\circ}$, the closure of $B_{1,1}^\circ$ in \mathbb{R}^2 , where $B_{1,1}^\circ$ is defined in (B.3). Thus, $\tilde{\Xi}_{1,1}$ can be considered as a proper subspace of $\overline{B_{1,1}^\circ}$ via embedding, which is denoted by $\tilde{\Xi}_{1,1} \hookrightarrow \overline{B_{1,1}^\circ}$. Note that $\tilde{\Xi}_{1,1}$ cannot be viewed as a subspace of $B_{1,1}^\circ$ or $B_{1,1}$, as $\delta = 0$ is allowed in $\tilde{\Xi}_{1,1}$ where the AR parameter ϕ and MA parameter θ are unidentified.

Andrews and Cheng (2012a) distinguishes the *true parameter space* from the *optimization parameter space* and assume that the true parameter space is contained in the interior of the optimization parameter space to exclude the *effects of boundary constraints*. Here, we also make this assumption. Based on the discussions mentioned above, one would think of defining the optimization parameter space for ξ as $\tilde{\Xi}_{1,1}$. However, $\tilde{\Xi}_{1,1}$ may not be a good choice for the optimization parameter space as it is not *compact* in \mathbb{R}^2 , which may lead to the situation that there is no maximum on $\tilde{\Xi}_{1,1}$. More importantly, there will be unit root effects when $|\phi|$ and $|\theta|$ approach to 1.

One way to define the optimization parameter space is as follows

$$\Xi_{1,1} \triangleq \{(\delta, \theta)' | \delta \in [\theta_- + \theta, \theta^- + \theta], \theta \in [\theta_-, \theta^-] \subset (-1, 1)\}, \quad (3.20)$$

which is compact in \mathbb{R}^2 . However, $\Xi_{1,1}$ thus defined is a little complicated to handle. For reasons that will be explained later, we choose to ignore the restraint on δ and use an *enlarged* optimization parameter space $\Xi_{1,1}^+$ which is defined as

$$\Xi_{1,1}^+ \triangleq \mathbb{R} \times \Theta, \quad (3.21)$$

where $\Theta \triangleq [\theta_-, \theta^-]$. It is obvious that $\Xi_{1,1}^+$ is closed in \mathbb{R}^2 , which will be sufficient for our purpose. The difference between adopting $\Xi_{1,1}^+$ and $\Xi_{1,1}$ is discussed in Corollary 3.2.19. For this study, we specify the true parameter space $\Xi_{1,1}^*$ as $\text{int}(\Xi_{1,1})$, the interior of $\Xi_{1,1}$. It is clear that

$$\Xi_{1,1}^* \triangleq \text{int}(\Xi_{1,1}) \subsetneq \Xi_{1,1} \subsetneq \tilde{\Xi}_{1,1} \hookrightarrow \overline{B_{1,1}^\circ}. \quad (3.22)$$

The asymptotic identification categories of drifting sequences of parameters are defined in the sense of identification by the asymptotic distributions of estimators. Suppose the asymptotic distribution $\pi_{\hat{\theta}_T}(\{\xi_T\}_{T=1}^\infty)$ exists for sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$. The sequence $\{\xi_T\}_{T=1}^\infty$ is said to fall into the *asymptotically unidentified* category for θ_T , if $\pi_{\hat{\theta}_T}(\{\xi_T\}_{T=1}^\infty)$ doesn't depend on sequence $\{\theta_T\}_{T=1}^\infty$, i.e., for any sequence $\{\tilde{\xi}_T^\diamond = (\delta_T^\diamond, \theta_T^\diamond)'\}_{T=1}^\infty$ with $\delta_T = \delta_T^\diamond$ for every $T \geq 1$,

$$\pi_{\hat{\theta}_T}(\{\tilde{\xi}_T^\diamond\}_{T=1}^\infty) = \pi_{\hat{\theta}_T}(\{\xi_T\}_{T=1}^\infty). \quad (3.23)$$

The sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ with asymptotic distribution $\pi_{\hat{\theta}_T}(\{\xi_T\}_{T=1}^\infty)$ is said to fall into the *asymptotically identified* category for θ_T , if it doesn't belong to the asymptotically unidentified category.

The asymptotically identified category for θ_T decomposes into three subcategories: weakly, semi-strongly, and strongly identified. A sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ in the asymptotically identified category is said to fall into the *weakly identified* subcategory, if the estimator $\hat{\theta}_T$ is not *consistent*, i.e., there exists $\epsilon > 0$, such that

$$\limsup_{T \rightarrow \infty} \Pr(|\hat{\theta}_T - \theta_T| > \epsilon) > 0. \quad (3.24)$$

For sequences in this subcategory, the finite-sample and asymptotic distributions of the estimator are in general *nonnormal*, and standard point estimates, hypothesis tests, and confidence intervals are usually unreliable. Stock, Wright and Yogo (2002) provide a survey of weak instruments and weak identification in generalized method of moments (GMM) and documents important features of weak identification. A sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ in the asymptotically identified category is said to fall into the *strongly identified* subcategory, if the estimator $\hat{\theta}_T$ has the two properties listed in (3.14); it is said to fall into the *semi-strongly identified* subcategory, if the estimator $\hat{\theta}_T$ has the two properties listed in (3.14), except that the convergence rate $T^{1/2}$ in (i) is replaced by some slower rate.

For the purpose of global analysis, we define the probability space for the ARMA(1,1) processes as follows. Assume that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is a probability space such that the identity function

$$\begin{aligned} Id_{\mathbb{R}} : (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ x &\mapsto x \end{aligned} \quad (3.25)$$

is a r.v. following the standard normal distribution, where $\mathcal{B}(X)$ denotes the *Borel σ -algebra* on topological space X . Define probability space $(\Omega, \mathcal{F}, \mathcal{P})$ as $\Omega \triangleq \mathbb{R}^{\mathbb{Z}}$, $\mathcal{F} \triangleq \mathcal{B}(\mathbb{R}^{\mathbb{Z}})$, and

$\mathcal{P} \triangleq \mu^{\mathbb{Z}}$, where $\mathbb{R}^{\mathbb{Z}}$ is equipped with *product topology*, such that

$$\begin{aligned} \epsilon_{t,T} : \quad (\Omega, \mathcal{F}, \mathcal{P}) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega \triangleq (\dots, \omega_{-1}, \omega_0, \omega_1, \dots)' &\mapsto \omega_t \end{aligned} \quad (3.26)$$

are i.i.d. r.v.'s inducing probability measure (p.m.) μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for a given $T \geq 1$ and every $t \in \mathbb{Z}$. Thus, innovation processes $\{\epsilon_{t,T}\}_{t=-\infty}^{\infty}$ are well defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ for all $T \geq 1$. As a result, ARMA(1,1) processes $\{Y_{t,T}\}_t$ are also well defined on $(\Omega, \mathcal{F}, \mathcal{P})$. Note that based on this simple construction of the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, we have

$$\epsilon_{t,T} = \epsilon_{t,T'} \quad (3.27)$$

for any $t \in \mathbb{Z}, T, T' \in \mathbb{Z}_+$.³¹

3.2.2 Quasi-Likelihood Function

We assume that $\{\epsilon_{t,T}\}_t \sim \text{i.i.d. } \mathcal{N}(0, 1)$, so the *Gaussian quasi-log-likelihood function* for $(\xi, \sigma)' = (\delta, \theta, \sigma)'$ conditional on $Y_{0,T}$ and $\epsilon_{0,t}$ multiplied by $-T^{-1}$ and ignoring a constant is given by (Andrews and Cheng (2012b) Formula (10.3))³²

$$Q_T(\xi, \sigma | \{Y_{t,T}\}_{t=0}^T) = \log(\sigma) + \frac{1}{2\sigma^2 T} \tilde{F}_T(\xi | \{Y_{t,T}\}_{t=0}^T), \quad (3.28)$$

³¹Note that if we only care about results like *(uniform) convergence in distribution* and *(uniform) convergence/boundedness in probability*, there is no need to define all the ARMA(p, q) processes $\{Y_{t,T}\}_t$ on the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$. However, if we want to study properties involving *(uniform) convergence a.s. (almost surely)*, and in L^p , we need to define all the ARMA(p, q) processes $\{Y_{t,T}\}_t$ on the same probability space and specify the relationship between $\{\epsilon_{t,T}\}_t$ and $\{\epsilon_{t,T'}\}_t$ when $T \neq T'$. In this situation, (3.27) is a natural treatment. Nevertheless, the results on (uniform) convergence in distribution and (uniform) convergence/boundedness in probability in this paper always hold no matter whether (3.27) holds and whether to define all the ARMA(p, q) processes $\{Y_{t,T}\}_t$ on the same probability space.

³²Note that function Q_T and Q_T^∞ below are actually well defined for all $\xi \in \mathbb{R} \times (-1, 1)$ a.e. (almost everywhere).

where

$$\tilde{F}_T(\xi|\{Y_{t,T}\}_{t=0}^T) \triangleq \sum_{t=1}^T \left(Y_{t,T} - \delta \sum_{i=0}^{t-1} (-\theta)^i Y_{t-i-1,T} \right)^2. \quad (3.29)$$

$Q_T(\xi, \sigma|\{Y_{t,T}\}_{t=0}^T)$ can be well approximated by $Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T)$ which is defined as

$$Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T) \triangleq \log(\sigma) + \frac{1}{2\sigma^2 T} \sum_{t=1}^T \left(Y_{t,T} - \delta \sum_{i=0}^{\infty} (-\theta)^i Y_{t-i-1,T} \right)^2. \quad (3.30)$$

The difference between $Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T)$ and $Q_T(\xi, \sigma|\{Y_{t,T}\}_{t=0}^T)$ lies in the *initial conditions*, which is asymptotically negligible (see discussion on *Initial Conditions Adjustment* in Andrews and Cheng (2012b)).

Let $\phi = \delta - \theta$, $Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T)$ can be rewritten as

$$Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T) = \log(\sigma) + \frac{1}{T} \sum_{t=1}^T \frac{[(1 + \theta L)^{-1}(1 - \phi L)Y_{t,T}]^2}{2\sigma^2}, \quad (3.31)$$

which is the log of the following (quasi-)likelihood function (multiplied by $-T^{-1}$ and ignoring a constant)

$$L_T(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{[(1 + \theta L)^{-1}(1 - \phi L)Y_{t,T}]^2}{2\sigma^2} \right). \quad (3.32)$$

We have the following (quasi-)log-likelihood function in terms of $(\xi, \sigma)'$

$$\mathcal{L}_T(\xi, \sigma) \triangleq Q_T^\infty(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T), \quad (3.33)$$

and the ARMA(1,1) model (3.1) is estimated by minimizing the criterion function $\mathcal{L}_T(\xi, \sigma)$ over $\Xi_{1,1}^+ \times (0, +\infty)$.³³

Note that for notational simplicity, we write $\mathcal{L}_T(\xi, \sigma)$ instead of $\mathcal{L}_T(\xi, \sigma|\{Y_{t,T}\}_{t=-\infty}^T)$.

³³Considering that we only observe $\{Y_{t,T}\}$ for $t = 1 - p, \dots, T$, \mathcal{L}_T would be called a QML (quasi maximum likelihood) criterion function. We prefer to call it an ML criterion function, as it is indeed the log likelihood function if $\{Y_{t,T}\}$ was known for $t \leq T$. For this reason, we call $(\hat{\xi}_T, \hat{\sigma}_T)'$ obtained by minimizing \mathcal{L}_T as the MLE in this paper. One form of the *exact* Gaussian likelihood function is given in (B.5).

As the observation $\{Y_{t,T}\}_{t=-\infty}^T$ is uniquely determined by sample point $\omega \in (\Omega, \mathcal{F}, \mathcal{P})$, $\mathcal{L}_T(\xi, \sigma)$ is essentially $\mathcal{L}_T(\xi, \sigma | \{Y_{t,T}(\omega)\}_{t=-\infty}^T)$. More precisely, \mathcal{L}_T is a *random function* as follows

$$\begin{aligned} \mathcal{L}_T : (\Omega, \mathcal{F}, \mathcal{P}) \times (\Xi_{1,1}^+ \times (0, +\infty)) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ (\omega, (\xi', \sigma)') &\mapsto \mathcal{L}_T(\xi, \sigma | \{Y_{t,T}(\omega)\}_{t=-\infty}^T) \end{aligned} \quad (3.34)$$

The function \mathcal{L}_T is measurable in its first argument and differentiable in the second. For notational simplicity, the dependence of \mathcal{L}_T on $\omega \in \Omega$ is frequently suppressed in this paper, and we often write $\mathcal{L}_T(\xi, \sigma)$ instead of $\mathcal{L}_T(\omega, (\xi, \sigma))$. Many other functions we have seen (e.g., Q_T , Q_T^∞ , and L_T) and those to be defined later (e.g., F_T , c_T , \mathcal{V}_T , C_T , and D_T , etc.) are all random functions treated in this way.

By (3.7) and (3.8), we have

$$\begin{aligned} &\sum_{t=1}^T [(1 + \theta L)^{-1} (1 - \phi L) Y_{t,T}]^2 \\ &= \sigma_T^2 \sum_{t=1}^T [(1 + \theta L)^{-1} (1 - \phi L) (1 - \phi_T L)^{-1} (1 + \theta_T L) \epsilon_{t,T}]^2. \end{aligned} \quad (3.35)$$

Define

$$F_T(\xi) \triangleq \sum_{t=1}^T [(1 + \theta L)^{-1} (1 - \phi L) (1 - \phi_T L)^{-1} (1 + \theta_T L) \epsilon_{t,T}]^2, \quad (3.36)$$

where $\phi = \delta - \theta$, then

$$\mathcal{L}_T(\xi, \sigma) = \log(\sigma) + \frac{\sigma_T^2}{2\sigma^2 T} F_T(\xi). \quad (3.37)$$

It is clear that finding $(\hat{\xi}_T, \hat{\sigma}_T)$ to minimize $\mathcal{L}_T(\xi, \sigma)$ is equivalent to finding $\hat{\xi}_T$ to minimize $F_T(\xi)$ first and then finding $\hat{\sigma}_T$ to minimize $\mathcal{L}_T(\hat{\xi}_T, \sigma)$.³⁴ This is why the parameter σ is actually irrelevant to the ML estimation of ξ which is the focus of this paper.

³⁴Note that the FOC $\frac{\partial \mathcal{L}_T}{\partial \sigma} = 0$ implies $\sigma^2 = \frac{\sigma_T^2 F_T}{T}$. An alternative treatment is to substitute this into (3.37), so that the minimization of \mathcal{L}_T is converted to minimization of the *reduced likelihood* $\frac{1}{2} \log(\frac{\sigma_T^2 F_T}{T})$.

Let $\xi_T = (\delta_T, \theta_T)'$, then it is clear that

$$F_T(\xi_T) = \sum_{t=1}^T \epsilon_{t,T}^2. \quad (3.38)$$

We have

$$F_T(\xi) = \sum_{t=1}^T [(1 - \delta L(1 + \theta L)^{-1})(1 + \delta_T L(1 - \phi_T L)^{-1})\epsilon_{t,T}]^2. \quad (3.39)$$

Let

$$A(r) = L(1 - rL)^{-1} = \sum_{j=0}^{\infty} r^j L^{j+1} \quad (3.40)$$

denote the corresponding (formal) power series in the lag operator L , and these power series are called *A-series*. Then

$$L(1 + \theta L)^{-1} = A(-\theta), L(1 - \phi_T L)^{-1} = A(\phi_T), \quad (3.41)$$

and

$$F_T = \sum [(1 - \delta A(-\theta))(1 + \delta_T A(\phi_T))\epsilon_{t,T}]^2, \quad (3.42)$$

where \sum and F_T are the abbreviations for notations $\sum_{t=1}^T$ and $F_T(\xi)$, respectively. Henceforth, we will use these abbreviations when there is no ambiguity. The expression in (3.42) is called the *A-form* of F_T associated with $(-\delta, -\theta, \delta_T, \phi_T)$. Expanding the *A-form* in (3.42) gives

$$F_T = \sum [(1 - \delta A(-\theta) + \delta_T A(\phi_T) - \delta \delta_T A(-\theta)A(\phi_T))\epsilon_{t,T}]^2. \quad (3.43)$$

3.2.3 A- and B- Series and Uniform Convergence/Boundedness in Probability

From Equation (3.43), we see that F_T is a linear combination of *stochastic series* of the form $\sum (B_1(\theta)\epsilon_{t,T})(B_2(\theta)\epsilon_{t,T})$, where $B_1(\theta)$ and $B_2(\theta)$ are nonrandom formal power series in the lag operator L (as functions of θ). These stochastic series are called *B-series*. To understand

the ML criterion function \mathcal{L}_T and the MLE, one first key step is to understand the behavior of B -series, the basic “building blocks” of F_T , over the entire parameter space Θ . Let’s start with A -series first.

Lemma 3.2.1 (Formulae of Derivatives).

$$A^{(j)}(r) = j!A^{j+1}(r), r \in \mathbb{R}, j \in \mathbb{N}, \quad (3.44)$$

where $A^{(j)}$ is the j -th (formal) derivative of A with respect to r , $A^{(0)} = A$.

Lemma 3.2.2 (Product-to-Sum Identity). For any $r, r' \in \mathbb{R}$,

$$(r' - r)A(r')A(r) = L[r'A(r') - rA(r)]. \quad (3.45)$$

We have the following convergence results about B -series $\sum(B_1\epsilon_{t,T})(B_2\epsilon_{t,T})$, where B_1 and B_2 are constant formal power series in L .

Lemma 3.2.3. Suppose the series $\sum_{j=0}^{\infty} q_j^2, \sum_{j=0}^{\infty} r_j^2 < \infty$, and let $Q = \sum_{j=0}^{\infty} q_j L^j$, and $R = \sum_{j=0}^{\infty} r_j L^j$.

Assume that $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim i.i.d.(0, \sigma^2)$ and $E\epsilon_t^4 = \sigma_4$.

1.

$$T^{-1/2} \left(\sum_{t=1}^T [\epsilon_t(Q\epsilon_t) - \sigma^2 q_0], \sum_{t=1}^T [\epsilon_t(R\epsilon_t) - \sigma^2 r_0] \right)' \xrightarrow{d} \mathcal{N}(0, V), \text{ as } T \rightarrow \infty, \quad (3.46)$$

where

$$V = \begin{pmatrix} (\sigma_4 - \sigma^4)q_0^2 + \sigma^4 \sum_{j=1}^{\infty} q_j^2 & (\sigma_4 - \sigma^4)q_0 r_0 + \sigma^4 \sum_{j=1}^{\infty} q_j r_j \\ (\sigma_4 - \sigma^4)q_0 r_0 + \sigma^4 \sum_{j=1}^{\infty} q_j r_j & (\sigma_4 - \sigma^4)r_0^2 + \sigma^4 \sum_{j=1}^{\infty} r_j^2 \end{pmatrix}. \quad (3.47)$$

2. Let $W_t = (Q\epsilon_t)(R\epsilon_t)$ for $t \in \mathbb{Z}$, then

$$T^{-1} \sum_{t=1}^T W_t \rightarrow \sigma^2 \sum_{j=0}^{\infty} q_j r_j \text{ a.s. and in } L^1, \text{ as } T \rightarrow \infty. \quad (3.48)$$

As a special case,

$$T^{-1} \sum_{t=1}^T \epsilon_t(R\epsilon_t) \rightarrow \sigma^2 r_0 \text{ a.s. and in } L^1, \text{ as } T \rightarrow \infty. \quad (3.49)$$

3. Let $\tilde{W}_t = W_t - E(W_t) = W_t - \sigma^2 \sum_{j=0}^{\infty} q_j r_j$ for $t \in \mathbb{Z}$, and denote the autocovariance function of $\{\tilde{W}_t\}_t$ by $\gamma(\cdot)$. Assume that $\sum_{j=0}^{\infty} |q_j|, \sum_{j=0}^{\infty} |r_j| < \infty$. Let $v = \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j)$, then $|v| < \infty$. If $v \neq 0$, then

$$T^{-1/2} \sum_{t=1}^T \tilde{W}_t \xrightarrow{d} \mathcal{N}(0, v), \text{ as } T \rightarrow \infty; \quad (3.50)$$

otherwise,

$$T^{-1/2} \sum_{t=1}^T \tilde{W}_t \xrightarrow{p} 0 \text{ and } \text{Var}(T^{-1/2} \sum_{t=1}^T \tilde{W}_t) \rightarrow 0, \text{ as } T \rightarrow \infty. \quad (3.51)$$

In Equation (3.43), (formal) power series $B_1(\theta)$ and $B_2(\theta)$ in $\sum (B_1(\theta)\epsilon_{t,T})(B_2(\theta)\epsilon_{t,T})$ are nonconstant functions of θ , and the convergence of these B -series as random functions is crucial to the global approach proposed in this paper. Lemma 3.2.3 has set up the *pointwise convergence/boundedness in probability* results for B -series, and we need some stronger results of *uniform convergence/boundedness in probability*. Roughly speaking, with *stochastic equicontinuity* condition, pointwise convergence can be strengthened to uniform convergence, which is a stochastic generalization of *Ascoli-Arzelà's theorem*. For results about stochastic equicontinuity and generic uniform convergence, see, Andrews (1987, 1992), Pollard (1990), Newey (1991), Pötscher and Prucha (1994), Billingsley (1999), etc.

In general, for an index set Θ (not necessarily equal to $[\theta_-, \theta^-]$), suppose that for each $T \in \mathbb{Z}_+$,

$$H_T : (\Omega, \mathcal{F}, \mathcal{P}) \times \Theta \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad (3.52)$$

is a random function which is measurable in its first argument. The sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$

converges to 0 in probability uniformly over Θ , as $T \rightarrow \infty$, if

$$\sup_{\theta \in \Theta} |H_T(\theta)| = o_p(1),^{35} \quad (3.53)$$

and is denoted by $H_T(\theta) = o_{p.u.}(1)$ or $H_T(\theta) \xrightarrow{p.u.} 0$, as $T \rightarrow \infty$. The sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is bounded in probability uniformly over Θ , as $T \rightarrow \infty$, if

$$\sup_{\theta \in \Theta} |H_T(\theta)| = O_p(1), \quad (3.54)$$

and is denoted by $H_T(\theta) = O_{p.u.}(1)$. The sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ converges to 0 a.s. uniformly over Θ , as $T \rightarrow \infty$, if

$$\sup_{\theta \in \Theta} |H_T(\theta)| \xrightarrow{a.s.} 0, \text{ as } T \rightarrow \infty, \quad (3.55)$$

and is denoted by $H_T(\theta) \xrightarrow{a.s.u.} 0$, as $T \rightarrow \infty$.

Now assume that Θ is a metric space with metric d_Θ , and let

$$B(\theta, r) \triangleq \text{closed ball of radius } r \text{ centered at } \theta \text{ in } (\Theta, d_\Theta). \quad (3.56)$$

The sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is *stochastically equicontinuous on Θ* , if for any $\epsilon > 0$, there exists $r > 0$ such that

$$\limsup_{T \rightarrow \infty} Pr \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, r)} |H_T(\theta') - H_T(\theta)| > \epsilon \right) < \epsilon. \quad (3.57)$$

For our case here, we want to have results about uniform boundedness instead of uniform convergence. For that purpose, let's define "stochastic equiboundedness" as follows.

³⁵In general, for operations like "sup" and "inf," etc., measurability could be an issue. One way to avoid this complications is to assume that probability statements like this are for *outer probability*. The outer probability of an arbitrary set $S \subset \Omega$ is defined as $\inf\{E(f) | f \text{ is measurable and } 1_S \leq f\}$. This treatment is adopted in Newey (1991). As all the random functions studied in this paper are defined in terms of smooth B -series and nonrandom functions, the sample paths of these random function are all smooth if we ignore a set of measure zero (or measure zero asymptotically). In this case, the measurability for operations like "sup" and "inf" is no longer an issue.

Definition 3.2.4. Given metric space (Θ, d_Θ) and sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ as in (3.52). The sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is stochastically equibounded on Θ if for any $\epsilon > 0$, there exists an $M > 0$ and $r > 0$ such that

$$\limsup_{T \rightarrow \infty} Pr \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, r)} |H_T(\theta') - H_T(\theta)| > M \right) < \epsilon. \quad (3.58)$$

It is clear that stochastic equicontinuity implies stochastic equiboundedness, and we have the following theorem which is parallel to Theorem 1 in Andrews (1992) (see Proposition B.1.3).

Theorem 3.2.5 (Generic Uniform Boundedness in Probability). Given metric space (Θ, d_Θ) and sequence $\{H_T\}_{T \in \mathbb{Z}_+}$ as in (3.52). For the following properties:

- (i) (Total Boundedness) (Θ, d_Θ) is a totally bounded metric space,
- (ii) (Pointwise Boundedness in Probability) $H_T(\theta) = O_p(1)$ for any $\theta \in \Theta$,
- (iii) (Stochastic Equiboundedness) The sequence $\{H_T\}_{T \in \mathbb{Z}_+}$ is stochastically equibounded on Θ ,
- (iv) (Uniform Boundedness in Probability) $H_T(\theta) = O_{p.u.}(1)$,

we have

1. (i), (ii), and (iii) \Rightarrow (iv),
2. (iv) \Rightarrow (ii) and (iii).

Lemma 3.2.6. Suppose $\Theta = [\theta_-, \theta^-]$, $Q(\theta) = \sum_{j=0}^{\infty} q_j(\theta) L^j$, and $R(\theta) = \sum_{j=0}^{\infty} r_j(\theta) L^j$, where $q_j(\theta)$ and $r_j(\theta)$ are continuous on Θ and differentiable on (θ_-, θ^-) for each $j \in \mathbb{N}$. Assume that $\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |q_j(\theta)| = M_q$, $\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |r_j(\theta)| = M_r$, $\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |q'_j(\theta)| = M'_q$, and $\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |r'_j(\theta)| = M'_r$ for every $\theta \in \Theta$, where $M_q, M_r, M'_q, M'_r < \infty$. Let $W_t(\theta) = (Q(\theta)\epsilon_t)(R(\theta)\epsilon_t)$

for $t \in \mathbb{Z}$, where $\{\epsilon_t\}_{t \in \mathbb{Z}} \sim \text{i.i.d.}(0, \sigma^2)$ and $E\epsilon_t^4 = \sigma_4$, then by Lemma 3.2.3, $E(W_t(\theta)) = \sigma^2 \sum_{j=0}^{\infty} q_j(\theta)r_j(\theta)$. Let $\tilde{W}_t(\theta) = W_t(\theta) - E(W_t(\theta)) = W_t - \sigma^2 \sum_{j=0}^{\infty} q_j(\theta)r_j(\theta)$ for $t \in \mathbb{Z}$, then

$$T^{-1/2} \sum_{t=1}^T \tilde{W}_t(\theta) = O_{p.u.}(1). \quad (3.59)$$

The above result can also be written as

$$\sum_{t=1}^T \tilde{W}_t(\theta) = O_{p.u.}(T^{1/2}). \quad (3.60)$$

An immediate corollary of the above lemma is as follows.

Corollary 3.2.7. *Given assumptions in Lemma 3.2.6,*

$$T^{-1} \sum_{t=1}^T W_t(\theta) = \sigma^2 \sum_{j \in \mathbb{N}} q_j(\theta)r_j(\theta) + o_{p.u.}(T^{(-1/2)+}) = O_{p.u.}(1).$$

We have the following lemma.

Lemma 3.2.8. *Let $\Psi_1(X_0, X_1, \dots, X_m)$ and $\Psi_2(Y_0, Y_1, \dots, Y_n)$ be two polynomials. Given any*

$\theta_1, \dots, \theta_m, \vartheta_1, \dots, \vartheta_n \in \Theta$, define

$$Q(\theta) = \Psi_1(A(-\theta), A(-\theta_1), \dots, A(-\theta_m)) \text{ and } R(\theta) = \Psi_2(A(-\theta), A(-\vartheta_1), \dots, A(-\vartheta_n)). \quad (3.61)$$

Then $Q(\theta)$ and $R(\theta)$ thus defined satisfy the conditions in Lemma 3.2.6, which implies that (3.60) is true.

3.2.4 First Order Conditions

The problem of minimizing \mathcal{L}_T will be checked by analyzing its first derivatives over optimization parameter space $\Xi_{1,1}^+$. Here we focus on the estimation of $\hat{\xi}_T$, which is the crucial part. Note that $F_T(\cdot, \theta)$ is not only a convex function of δ but a *parabola*, which can also be

seen from the second partial derivative of F_T shown below. We have

$$\frac{\partial F_T}{\partial \delta} = 2 \sum [(1 - \delta A(-\theta))(1 + \delta_T A(\phi_T))\epsilon_{t,T}](-P(\theta, \delta_T, \phi_T)\epsilon_{t,T}), \quad (3.62)$$

where

$$P(\theta, \delta_T, \phi_T) = A(-\theta)(1 + \delta_T A(\phi_T)), \quad (3.63)$$

and

$$\frac{\partial^2 F_T}{\partial \delta^2} = 2 \sum (P\epsilon_{t,T})^2 \geq 0, \quad (3.64)$$

which is a constant (when θ and ϕ_T are fixed). An immediate corollary is that $\frac{\partial F_T}{\partial \delta} = 0$ is a necessary and sufficient condition for finding the minimum of $F_T(\cdot, \theta)$ over \mathbb{R} , when θ is given. (Unfortunately, $F_T(\delta, \cdot)$ is not a convex function of θ . Otherwise, the problem could be much easier!)

The boundary of $\Xi_{1,1}^+$ in \mathbb{R}^2 is

$$\partial \Xi_{1,1}^+ = \mathbb{R} \times \{\theta_-, \theta^-\}. \quad (3.65)$$

It is clear that

$$\frac{\partial \mathcal{L}_T}{\partial \theta} = \frac{\sigma_T^2}{2\sigma^2 T} \frac{\partial F_T}{\partial \theta} \text{ and } \frac{\partial \mathcal{L}_T}{\partial \delta} = \frac{\sigma_T^2}{2\sigma^2 T} \frac{\partial F_T}{\partial \delta}. \quad (3.66)$$

If $\hat{\xi}_T \in \text{int}(\Xi_{1,1}^+)$, i.e., $\hat{\theta}_T \in (\theta_-, \theta^-)$, then it is the case that

$$\frac{\partial F_T}{\partial \theta}(\hat{\xi}_T) = 0 \text{ and } \frac{\partial F_T}{\partial \delta}(\hat{\xi}_T) = 0; \quad (3.67)$$

otherwise, we have $\hat{\xi}_T \in \partial \Xi_{1,1}^+$, i.e., $\hat{\theta}_T \in \{\theta_-, \theta^-\}$, and it is the case that

$$\frac{\partial F_T}{\partial \delta}(\hat{\xi}_T) = 0. \quad (3.68)$$

So the advantages of adopting the enlarged optimization parameter space $\Xi_{1,1}^+$ are that the

boundary of $\Xi_{1,1}^+$ is very simple, and the FOC $\frac{\partial F_T}{\partial \delta} = 0$ always holds at $\xi = \hat{\xi}_T$.

Based on above discussion, since we choose $\Xi_{1,1}^+$ as the optimization parameter space, the problem of minimizing F_T over $\Xi_{1,1}^+$ can be implemented as follows. For each $\theta \in [\theta_-, \theta^-]$, define

$$f_T(\theta) = \arg \min_{\delta \in \mathbb{R}} F_T(\delta, \theta).^{36} \quad (3.69)$$

We know that $f_T(\theta)$ is the unique value of δ that minimizes $F_T(\cdot, \theta)$, i.e., the trough of the parabola. We thus obtained a curve

$$\begin{aligned} c_T : \Theta &\rightarrow \Xi_{1,1}^+ \\ \theta &\mapsto (f_T(\theta), \theta)' \end{aligned} \quad (3.70)$$

in $\Xi_{1,1}^+$ which is actually the valleys of F_T , and the global minimum of F_T over $\Xi_{1,1}^+$ must sit somewhere on curve c_T . Define

$$\begin{aligned} \mathcal{V}_T : \Theta &\rightarrow \Xi_{1,1}^+ \\ \theta &\mapsto F_T(c_T(\theta)) \end{aligned}, \quad (3.71)$$

then $\mathcal{V}_T'(\hat{\theta}_T) = 0$ if the minimum of \mathcal{V}_T occurs in (θ_-, θ^-) . Furthermore, we have

$$\mathcal{V}_T'(\theta) = \frac{\partial F_T}{\partial \theta} + \frac{\partial F_T}{\partial \delta} \frac{df_T}{d\theta} = \frac{\partial F_T}{\partial \theta}(c_T(\theta)). \quad (3.72)$$

By Lemma 3.2.8, we have³⁷

Lemma 3.2.9. *Given a drifting sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, the FOC $\frac{\partial \mathcal{L}_T}{\partial \delta} = 0$ gives the following*

³⁶The dependence of f_T on ω , δ_T , and ϕ_T (or θ_T) are suppressed here, like many other functions introduced in this paper. We will not mention facts like this hereon, unless there might be possible ambiguities.

³⁷The $O(T^a)$ terms in this paper are sequences of constants that may depend on parameter values, i.e., they are *nonrandom* piecewise continuous functions of δ , θ , etc. Terms like $O_p(T^a)$ and $o_{p.u.}(T^a)$, etc., occurred below are sequences of r.v.'s that may depend on parameter values, i.e., they are random functions piecewise continuous in δ , θ , etc., *a.e.* More specifically, given $T \geq 1$, for almost every $\omega \in (\Omega, \mathcal{F}, \mathcal{P})$, the sample path of the random function associated with ω is piecewise continuous in those parameters.

equation on Θ

$$\delta(\theta, \delta_T, \phi_T) = T^{-1/2}(1 - \theta^2)C_T(\theta) + \delta_T \left(\frac{1 - \theta^2}{1 + \phi_T \theta} \frac{1 + \phi_T \theta_T}{1 - \phi_T^2} \right) f_1(\theta, \delta_T, \phi_T) + \varepsilon_{1,T}, \quad (3.73)$$

where $\varepsilon_{1,T} = O_{p.u.}(T^{-1/2} \max\{T^{-1/2}, |\delta_T|\})$,

$$C_T(\theta) \triangleq T^{-1/2} \sum \epsilon_{t,T}(A(-\theta)\epsilon_{t,T}),^{38} \quad (3.74)$$

and

$$f_1(\theta, \delta_T, \phi_T) = \left[1 - \delta_T \frac{2\theta}{1 + \phi_T \theta} + \delta_T^2 \frac{1 - \phi_T \theta}{(1 + \phi_T \theta)(1 - \phi_T^2)} \right]^{-1} = 1 + O(|\delta_T|). \quad (3.75)$$

Thus, we have

$$f_T = O_{p.u.}(\max\{T^{-1/2}, |\delta_T|\}). \quad (3.76)$$

As a corollary,

$$\hat{\delta}_T = O_p(\max\{T^{-1/2}, |\delta_T|\}). \quad (3.77)$$

Remark 3.2.10. The above lemma doesn't require that the drifting sequence $\{\xi_T\}_{T=1}^\infty$ be convergent, which is also the case for Theorem 3.2.11, Lemma 3.2.12, Lemma 3.2.14, and Theorem 3.3.1.

By Lemma 3.2.3,

$$C_T(\theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{1 - \theta^2}) \text{ for any } \theta \in (-1, 1), \text{ as } T \rightarrow \infty. \quad (3.78)$$

Similar to the case of \mathcal{L}_T , we have a random function

$$\begin{aligned} C_T : (\Omega, \mathcal{F}, \mathcal{P}) \times (-1, 1) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ (\omega, \theta) &\mapsto T^{-1/2} \sum \epsilon_{t,T}(\omega) [A(-\theta)\epsilon_{t,T}(\omega)] \end{aligned} \quad (3.79)$$

³⁸Rigorously speaking, C_T is not well defined on a set of measure zero where the series $\epsilon_{t,T}(A(-\theta)\epsilon_{t,T})$ diverge, and we simply ignore this null set. For all the B -series and random functions constructed from B -series in this paper, we always ignore those null sets.

When the dependence of C_T on $\omega \in \Omega$ is suppressed, $C_T(\theta)$ also denotes a r.v. on $(\Omega, \mathcal{F}, \mathcal{P})$ for each $\theta \in (-1, 1)$. Given $\omega \in (\Omega, \mathcal{F}, \mathcal{P})$, define

$$\begin{aligned} C_{\omega, T} : (-1, 1) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \theta &\mapsto C_T(\omega, \theta) \end{aligned}, \quad (3.80)$$

and we call the function $C_{\omega, T}$ of θ the *sample path* of C_T associated with ω . Note that the sample path $C_{\omega, T}$ is an infinitely differentiable function of θ in the space $(\Omega, \mathcal{F}, \mathcal{P})$ for any $T \in \mathbb{Z}_+$.

We have the following theorem.

Theorem 3.2.11 (Consistency of $\hat{\delta}_T$). *Given a drifting sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, we have*

$$f_T(\theta) - \delta_T = \frac{f_1(\theta, \delta_T, \phi_T)}{(1 + \phi_T \theta)(1 - \phi_T^2)} f_2(\theta, \delta_T, \phi_T) (\theta - \theta_T) \delta_T + T^{-1/2} (1 - \theta^2) C_T(\theta) + \varepsilon_{1,T},$$

where f_1 and $\varepsilon_{1,T}$ are defined in Lemma 3.2.9,

$$f_2(\theta, \delta_T, \phi_T) = (1 - \phi_T^2) \phi_T - (1 - \phi_T^2) (\theta - \theta_T) + \delta_T \phi_T (-\theta + \phi_T). \quad (3.81)$$

As a corollary, we have

$$\begin{aligned} \Delta \hat{\delta}_T &= \frac{f_1(\hat{\theta}_T, \delta_T, \phi_T)}{(1 + \phi_T \hat{\theta}_T)(1 - \phi_T^2)} f_2(\hat{\theta}_T, \delta_T, \phi_T) \Delta \hat{\theta}_T \delta_T + T^{-1/2} (1 - \hat{\theta}_T^2) C_T(\hat{\theta}_T) + \varepsilon_{2,T} \\ &= O_p(\max\{T^{-1/2}, |\Delta \hat{\theta}_T \delta_T|\}), \end{aligned}$$

where

$$\varepsilon_{2,T} = O_p(T^{-1/2} \max\{T^{-1/2}, |\delta_T|\}). \quad (3.82)$$

Now let's look at $\frac{\partial F_T}{\partial \theta}$. We have

$$\begin{aligned} \frac{\partial F_T}{\partial \theta} &= 2 \sum [(1 - \delta A(-\theta))(1 + \delta_T A(\phi_T)) \epsilon_{t,T}] [\delta A(-\theta)^2 (1 + \delta_T A(\phi_T)) \epsilon_{t,T}] \\ &= 2 \delta \sum [(1 - \delta A(-\theta) + \delta_T A(\phi_T) - \delta \delta_T A(-\theta) A(\phi_T)) \epsilon_{t,T}] [A(-\theta)^2 (1 + \delta_T A(\phi_T)) \epsilon_{t,T}]. \end{aligned} \quad (3.83)$$

It's clear that F_T becomes flat in the direction of θ when δ is close to 0, which may lead to numerical inaccuracy in practice. We have the following lemma.

Lemma 3.2.12. *Given a drifting sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, we have*

$$\begin{aligned} & \frac{\partial F_T}{\partial \theta} \\ = & 2\delta T^{1/2} \left\{ D_T(\theta) + T^{1/2} \left[\frac{\delta\theta}{(1-\theta^2)^2} + \delta_T \left(\frac{\phi_T}{(1+\phi_T\theta)^2} + f_3(\xi, \delta_T, \phi_T) \right) \right] + O_{p,u}(\max\{|\delta|, |\delta_T|\}) \right\}, \end{aligned} \quad (3.84)$$

where

$$f_3(\xi, \delta_T, \phi_T) = \frac{-\delta\theta^2}{(1+\phi_T\theta)(1-\theta^2)^2} + \frac{\delta_T\phi_T^2}{(1+\phi_T\theta)^2(1-\phi_T^2)} - \frac{\delta(1+\phi_T\theta^3)}{(1-\theta^2)^2(1+\phi_T\theta)^2} - \frac{\delta\delta_T(\phi_T-\theta-\phi_T\theta^2+\phi_T^2\theta^3)}{(1-\phi_T^2)(1-\theta^2)^2(1+\phi_T\theta)^2}, \quad (3.85)$$

and

$$D_T(\theta) \triangleq T^{-1/2} \sum \epsilon_{t,T}(A(-\theta)^2 \epsilon_{t,T}). \quad (3.86)$$

Remark 3.2.13. *Let $f_4(\xi, \delta_T, \phi_T) \triangleq \frac{\delta\theta}{(1-\theta^2)^2} + \delta_T \left(\frac{\phi_T}{(1+\phi_T\theta)^2} + f_3(\xi, \delta_T, \phi_T) \right)$, then*

$$E \left([(1 - \delta A(-\theta))(1 + \delta_T A(\phi_T)) \epsilon_{t,T}] [\delta A(-\theta)^2 (1 + \delta_T A(\phi_T)) \epsilon_{t,T}] \right) = f_4(\xi, \delta_T, \phi_T).$$

By Lemma 3.2.3,

$$D_T(\theta) = T^{-1/2} \sum_{j=2}^{\infty} (j-1)(-\theta)^{j-2} \epsilon_{t-j,T} \epsilon_{t,T} \xrightarrow{d} \mathcal{N}(0, v(\theta)) \text{ for } \theta \in (-1, 1), \text{ as } T \rightarrow \infty, \quad (3.87)$$

where

$$v(\theta) = \sum_{j=1}^{\infty} (j(-\theta)^{j-1})^2 = \frac{1+\theta^2}{(1-\theta^2)^3}. \quad (3.88)$$

Similar to the case of C_T , we can define D_T as

$$\begin{aligned} D_T : (\Omega, \mathcal{F}, \mathcal{P}) \times (-1, 1) & \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ (\omega, \theta) & \mapsto T^{-1/2} \sum \epsilon_{t,T}(\omega) [A^2(-\theta) \epsilon_{t,T}(\omega)] \end{aligned}, \quad (3.89)$$

and then define $D_{\omega,T}$ accordingly. An interesting and important fact to mention is that, by (3.74), (3.86), and Lemma 3.2.1,

$$D_{\omega,T}(\theta) = -C'_{\omega,T}(\theta) \quad (3.90)$$

for any $\theta \in (-1, 1)$ on $(\Omega, \mathcal{F}, \mathcal{P})$ a.e.³⁹

Lemma 3.2.14. *Given a drifting sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, we have*

$$\mathcal{V}'_T(\theta) = \frac{\partial F_T}{\partial \theta}(f_T(\theta), \theta) = 2\mathcal{W}_{1,T}(\theta)\mathcal{W}_{2,T}(\theta), \quad (3.91)$$

where

$$\mathcal{W}_{1,T}(\theta) = T^{1/2}f_T(\theta) = (1 - \theta^2)C_T(\theta) + d_T \left(\frac{1 - \theta^2}{1 + \phi_T\theta} \frac{1 + \phi_T\theta_T}{1 - \phi_T^2} \right) f_1(\theta, \delta_T, \phi_T) + T^{1/2}\varepsilon_{1,T}, \quad (3.92)$$

$$\begin{aligned} \mathcal{W}_{2,T}(\theta) &= D_T(\theta) + \frac{\theta C_T(\theta)}{1 - \theta^2} + \frac{T^{1/2}\theta\varepsilon_{1,T}}{1 - \theta^2} + O_{p.u}(\max\{|f_T(\theta)|, |\delta_T|\}) \\ &\quad + \frac{d_T}{1 + \phi_T\theta} \left(\frac{\theta}{1 - \theta^2} \frac{1 + \phi_T\theta_T}{1 - \phi_T^2} f_1(\theta, \delta_T, \phi_T) + \frac{\phi_T}{1 + \phi_T\theta} + (1 + \phi_T\theta)f_3(c_T(\theta), \delta_T, \phi_T) \right), \end{aligned} \quad (3.93)$$

$$d_T = T^{1/2}\delta_T, \quad (3.94)$$

$f_T, f_1, \varepsilon_{1,T}$, and f_3 are defined in (3.69), Lemma 3.2.9, and Lemma 3.2.12, respectively.

Remark 3.2.15. *Let*

$$f_5(\xi, \delta_T, \phi_T) \triangleq \frac{1}{1 + \phi_T\theta} \frac{\theta}{1 - \theta^2} \frac{1 + \phi_T\theta_T}{1 - \phi_T^2} f_1(\theta, \delta_T, \phi_T) + \frac{\phi_T}{(1 + \phi_T\theta)^2} + f_3(\xi, \delta_T, \phi_T), \quad (3.95)$$

then $f_5(\xi_T, \delta_T, \phi_T) = 0$ and $\frac{\partial f_5}{\partial \theta}(\xi_T, \delta_T, \phi_T) = \frac{1 - \phi_T^2 + \delta_T[2\phi_T(1 + \phi_T^2) - \delta_T(1 + 3\phi_T^2) + \delta_T^2\phi_T]}{(1 + \phi_T\theta_T)(1 - \phi_T^2)(1 - \theta_T^2)^2}$.

³⁹The only situations that $C_{\omega,\infty}$ fails to be infinitely differentiable with respect to θ is when $C_\infty(\theta)$ and its derivatives are not *well defined*, i.e., the series (3.74) and its (formal) derivatives don't converge. It is clear that this *exception set* is of measure zero.

3.2.5 Gaussian Processes and Convergence in Finite-Dimensional Distributions

By Lemma 3.2.3, we have

$$(C_T(\theta), D_T(\theta))' \xrightarrow{d} \mathcal{N}(0, V_{C,D})(\theta)) \text{ for any } \theta \in (-1, 1), \text{ as } T \rightarrow \infty, \quad (3.96)$$

where

$$V_{C,D}(\theta) = \begin{pmatrix} \frac{1}{1-\theta^2} & -\frac{\theta}{(1-\theta^2)^2} \\ -\frac{\theta}{(1-\theta^2)^2} & \frac{1+\theta^2}{(1-\theta^2)^3} \end{pmatrix} = \text{Cov}((C_T(\theta), D_T(\theta))'), \quad (3.97)$$

and

$$\det(V_{C,D}(\theta)) = \frac{1}{(1-\theta^2)^4} \geq 1. \quad (3.98)$$

The pointwise weak convergence results in (3.78), (3.87) and (3.96) tells us that the random vector $X_T(\theta) \triangleq (C_T(\theta), D_T(\theta))'$ is asymptotically a *Gaussian process (GP)*, i.e.,

$$X_T(\theta) \xrightarrow{d} \mathcal{GP}(m(\theta), k(\theta, \theta')), \text{ as } T \rightarrow \infty, \quad (3.99)$$

where $m(\theta) = E[X_T(\theta)] = 0$ and $k(\theta) = E[X_T(\theta)X_T(\theta)']$ for $\theta \in (-1, 1)$. For more discussions about GP's, see, e.g., Rasmussen and Williams (2006). However, it is not enough to treat $X_T(\theta)$ and random functions like this as GP's for our purpose. A related notion of *weak convergence to a family of distributions* is given in Mikusheva (2007) and Gorodnichenko, Mikusheva and Ng (2012), which is in some sense stronger than the notion of converging to a GP.

Here, as the global approach is based on analyzing the sample paths of $\mathcal{V}_T(\theta)$, we adopt a another notion of weak convergence which is very natural when random functions C_T and D_T are viewed as *stochastic processes*.

Definition 3.2.16. Given a sequence of random functions $\{\tilde{X}_T\}_{T=1}^\infty$ and random function \tilde{X}_∞ ,

where $\tilde{X}_T : (\Omega_T, \mathcal{F}_T, \mathcal{P}_T) \times \Theta \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and $\tilde{X}_\infty : (\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty) \times \Theta \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for every $T \in \mathbb{Z}_+$ and some internal $\Theta \subset \mathbb{R}$, and $(\Omega_T, \mathcal{F}_T, \mathcal{P}_T)$, $(\Omega_{T'}, \mathcal{F}_{T'}, \mathcal{P}_{T'})$, and $(\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty)$ can be mutually distinct when $T \neq T'$. The sequence $\{\tilde{X}_T\}_T$ is said to converge in finite-dimensional distributions to \tilde{X}_∞ (written $\tilde{X}_T \xrightarrow{f.d.d.} \tilde{X}_\infty$) if for any $m \in \mathbb{Z}_+$, $\theta_1 < \theta_2 < \dots < \theta_m \in \Theta$, and $S \in \mathcal{B}(\mathbb{R}^{mn})$,

$$Pr[(\tilde{X}_T(\theta_1), \dots, \tilde{X}_T(\theta_m))' \in S] \rightarrow Pr[(\tilde{X}_\infty(\theta_1), \dots, \tilde{X}_\infty(\theta_m))' \in S], \text{ as } T \rightarrow \infty. \quad (3.100)$$

More generally, the term $Pr[(\tilde{X}_\infty(\theta_1), \dots, \tilde{X}_\infty(\theta_m))' \in S]$ in (3.100) can be replaced with $\mathcal{Q}_{\underline{\theta}}(S)$, where $\underline{\theta} = (\theta_1, \dots, \theta_m)$, $\mathcal{Q}_{\underline{\theta}}$ is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, the collection $\{\mathcal{Q}_{\underline{\theta}}\}_{\underline{\theta}}$ is a *consistent family of finite-dimensional distributions*. For details about “consistent family of finite-dimensional distributions,” see, for instance, Karatzas and Shreve (1991, Chap. 2).

Define $(\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty)$ as follows: $\Omega_\infty \triangleq \mathbb{R}^{\mathbb{Z}_+}$, $\mathcal{P}_\infty \triangleq \mathcal{B}(\mathbb{R}^{\mathbb{Z}_+})$, and $\mathcal{P}_\infty \triangleq \mu^{\mathbb{Z}_+}$, where $\mathbb{R}^{\mathbb{Z}_+}$ is equipped with product topology and μ is defined in (3.25), such that

$$\begin{aligned} \eta_m : (\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega &\triangleq (\omega_1, \omega_2, \dots)' \mapsto \omega_m \end{aligned} \quad (3.101)$$

are i.i.d. r.v.'s inducing p.m. μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for all $m \in \mathbb{Z}_+$. Let

$$C_\infty(\theta) \triangleq \sum_{m=1}^{\infty} (-\theta)^m \eta_m \text{ and } D_\infty(\theta) \triangleq \sum_{m=2}^{\infty} m(-\theta)^{m-1} \eta_m, \quad (3.102)$$

then random functions $C_\infty(\theta)$ and $D_\infty(\theta)$ are well defined on $(\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty)$ a.e., and

$$D_{\omega, \infty}(\theta) = -C'_{\omega, \infty}(\theta) \quad (3.103)$$

holds for any $\theta \in (-1, 1)$ on $(\Omega_\infty, \mathcal{F}_\infty, \mathcal{P}_\infty)$ a.e.

By CLT (central limit theorem), we have $T^{-1/2} \sum \epsilon_{t,T} \epsilon_{t-m,T} \xrightarrow{d} \eta_m$ as $T \rightarrow \infty$ for any $m \in \mathbb{Z}_+$. By this fact and Cramér-Wold device, it is clear that

$$X_T \xrightarrow{f.d.d.} X_\infty \triangleq (C_\infty, D_\infty)', \text{ as } T \rightarrow \infty. \quad (3.104)$$

As mentioned earlier, the purpose of examining the notion of convergence in finite-dimensional distributions is to further analyze the sample paths of $\mathcal{V}_T(\theta)$. For a.e. (almost every) $\omega \in S_T^\delta(\theta_-, \theta^-, \xi_T) \subset (\Omega, \mathcal{F}, \mathcal{P})$, where $S_T^\delta(\theta_-, \theta^-, \xi_T)$ is defined in Footnote 44 whose probability approaches 1 as $T \rightarrow \infty$, the sample path $\mathcal{V}_{\omega,T}(\theta)$ associated with ω is a smooth function of θ on Θ , i.e., $\mathcal{V}_{\omega,T}(\theta)$ is of class C^∞ , or equivalently, $\mathcal{V}_{\omega,T}(\theta) \in C^\infty(\Theta)$.

3.2.6 Asymptotic Distributions

It has been shown that some terms in (3.91) are higher-order infinitesimals, and for different sequence $\{\xi_T\}_T$, they will be treated accordingly.

Definition 3.2.17. *Let*

$$h(\theta|\phi) \triangleq \frac{1}{1+\phi\theta} \left(\frac{\theta}{1-\theta^2} + \frac{\phi}{1+\phi\theta} \right), \quad (3.105)$$

and

$$G_T(\theta) \triangleq D_T(\theta) + \frac{\theta C_T(\theta)}{1-\theta^2}. \quad (3.106)$$

The (random) function

$$\mathcal{S}_T(\theta|d_T, \phi_T) \triangleq 2 \left[(1-\theta^2)C_T(\theta) + \frac{(1-\theta^2)d_T}{1+\phi_T\theta} \right] [G_T(\theta) + d_T h(\theta|\phi_T)] \quad (3.107)$$

is called the shape analysis function with sample size T , where d_T is defined in (3.94).

By (3.96)–(3.98), we know that the GP

$$G_T(\theta) \xrightarrow{d} \mathcal{N}(0, (1 - \theta^2)^{-3}), \text{ as } T \rightarrow \infty \text{ for every } \theta \in (-1, 1). \quad (3.108)$$

By Formula (3.97), $\frac{\theta}{1-\theta^2} = -\frac{\text{Cov}(C_T(\theta), D_T(\theta))}{\text{Var}(C_T(\theta))}$, which implies that

$$E(C_T(\theta)G_T(\theta)) = E\left(C_T(\theta)D_T(\theta) - C_T(\theta)^2 \frac{\text{Cov}(C_T(\theta), D_T(\theta))}{\text{Var}(C_T(\theta))}\right) = 0. \quad (3.109)$$

By (3.96)–(3.98), we have

$$(C_T(\theta), G_T(\theta))' \xrightarrow{d} \mathcal{N}(0, V_{C,G}(\theta)) \text{ for any } \theta \in (-1, 1), \text{ as } T \rightarrow \infty, \quad (3.110)$$

where

$$V_{C,G}(\theta) = \begin{pmatrix} \frac{1}{1-\theta^2} & 0 \\ 0 & \frac{1}{(1-\theta^2)^3} \end{pmatrix} = \text{Cov}((C_T(\theta), G_T(\theta))'), \quad (3.111)$$

and

$$\det(V_{C,G}(\theta)) = \frac{1}{(1-\theta^2)^4} = \det(V_{C,D}(\theta)) \geq 1. \quad (3.112)$$

Lemma 3.2.14 and the shape analysis function are the workhorse for the global approach. Before analyzing the asymptotic distributions in the root cancellation case, let's recall the classical result on asymptotic distributions of MLE in the usual ARMA(p, q) models (i.e., strongly identified non root-cancellation case). We have the classical results such as Proposition B.1.4 and Corollary B.1.5. From (B.12) in Corollary B.1.5, we see that the magnitude of δ^* doesn't "really" affect the variance of $T^{1/2}(\tilde{\delta}_T - \delta^*)$, which coincides with the result on ROC of $\triangle \hat{\delta}_T$ in Table 3.1.

We have Theorem 3.1.2.

Remark 3.2.18. By Formula (3.109), we have

$$E\mathcal{S}_T(\theta|0, \phi_T) = 2(1 - \theta^2)E(C_T(\theta)G_T(\theta)) = 0, \quad (3.113)$$

which in some sense “coincides” with the fact that θ_T is asymptotically unidentified in Case (1) of Theorem 3.1.2. In this case, the asymptotic distribution $\pi_{\hat{\theta}_T}$ spreads out over the entire region $[\theta_-, \theta^-]$. Another interesting fact is that $\mathcal{V}(\theta_0) = [V_{C,D}(\theta_0)]^{-1}$, where $V_{C,D}(\theta)$ is defined in (3.97).

By (B.12), (3.6), and (B.101), we see that the asymptotic distribution in the semi-strongly identified case “matches” that in the strongly identified case.

Corollary 3.2.19 (Irrelevance of Optimization Parameter Space). *Given a $1/2\text{-}\langle\infty\rangle$ sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ and spaces $\Xi^{(i)}$ ($i = 1, 2$) whose interior contains $\{\xi_T\}_{T=1}^\infty$, then the MLE $\hat{\xi}_T^{(i)}$ obtained from adopting the optimization parameter space $\Xi^{(i)}$ are asymptotically the same, i.e., $\lim_{T \rightarrow \infty} \Pr(\hat{\xi}_T^{(1)} = \hat{\xi}_T^{(2)}) = 1$.*

Note that in no and weak identification cases, the choice of the optimization parameter space does affect the (asymptotic) distributions of $\hat{\xi}_T$. The simulation results do not support the statement in Andrews and Cheng (2012a, p. 2166): “Given $\text{int}(\Theta) \supset \Theta^*$, the true value of θ cannot lie on the boundary of the optimization parameter space. In consequence, the asymptotic distribution of $\hat{\theta}_n$ is not affected by boundary constraints for any sequence of true parameters in Θ^* .” In the case of weak and no identification, the choice of optimization parameter space has a profound effect on the distribution of $\hat{\theta}_T$. Figure 3.1 depicts the simulation results of MLE $\hat{\theta}_T$ for different choices of Θ , which are shown in different colors.

We have Theorem 3.1.3, and its proof is straightforward by previous results.

Theorem 3.1.2 and Theorem 3.1.3 together provide a complete classification of drifting sequences of true distributions in terms of asymptotic distributions. Namely, a drifting se-

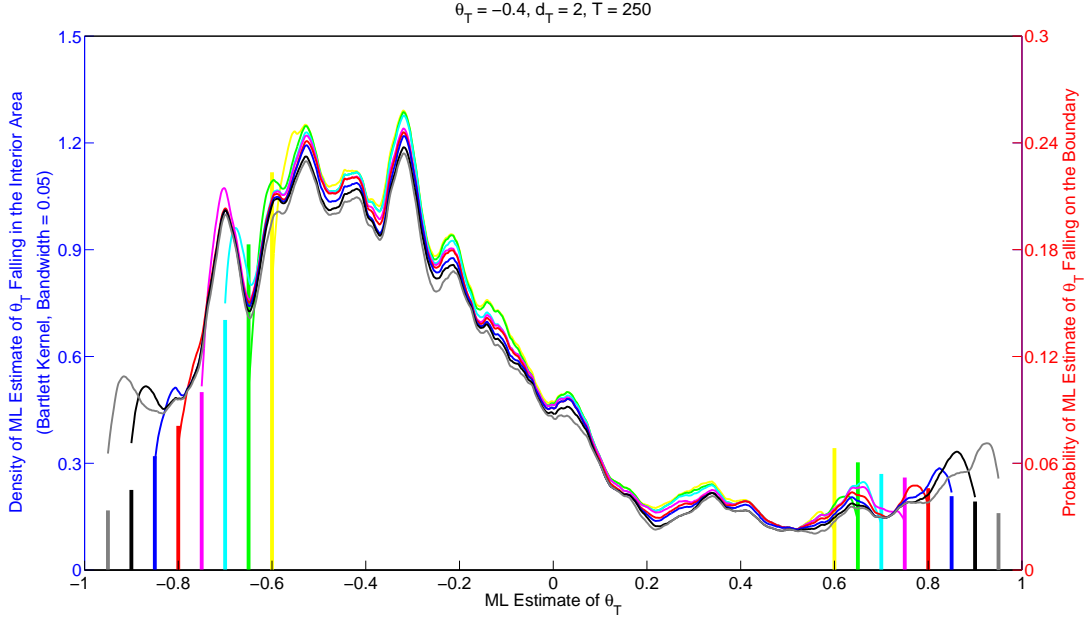


Figure 3.1: Boundary Effects on $\hat{\theta}_T$

Notes: Eight different choices of $\Theta = [-\theta^-, \theta^-]$ are illustrated in this figure, where $\theta^- = 0.60, 0.65, \dots, 0.90$ and 0.95 , $\theta_T = -0.4$, $d_T = 2$, and $T = 250$. Given a choice of Θ , the vertical bars at $\theta = \pm\theta^-$ indicates the probability that $\hat{\theta}_T$ falls on the boundary $\partial\Theta(= \{-\theta^-, \theta^-\})$. The corresponding curve, obtained by using the Bartlett kernel with a bandwidth of 0.05 from $100,000$ simulations, is the kernel estimation of the density of $\hat{\theta}_T$ over $\text{int}(\Theta)(= (-\theta^-, \theta^-))$.

quence of true distributions can be classified as (1) non, (2) weakly, (3) semi-strongly, or (4) strongly identified, or (5) having no asymptotic distribution. For all these five cases, Theorem 3.1.2 and Theorem 3.1.3 provide corresponding necessary and sufficient conditions.

We have Theorem 3.1.4 which immediately follows the above results.

Corollary 3.2.20 ($T^{1/2}$ -Consistency of $\hat{\delta}_T$). *Given a drifting sequence $\{\tilde{\zeta}_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$ whose asymptotic distribution exists, we have $\Delta\hat{\delta}_T = O_p(T^{-1/2})$.*

3.3 Hypothesis Testing

Testing hypothesis in the presence of weak identification has attracted attention of many econometricians. For the root cancellation case here, various test are studied in the literature mentioned previously, for example, Andrews and Cheng (2012a) explore t and QLR tests and confidence sets. Andrews and Mikusheva (2012) propose a particular form of the classical LM (Lagrange multiplier) tests (LM_o and LM_e statistic) for testing simple hypotheses on the full parameter vector, which are robust to weak identification and cover a large number of cases including the ARMA(1,1) with root cancellation case here.⁴⁰ A closely related LM test is suggested by Qu (2011) for a certain type of log-linearized DSGE models.

Here for the ARMA(1,1) model, we propose a statistic, which we refer to as the J -statistic, for testing a simple hypothesis on the parameter vector ζ , and the corresponding test is robust to the full range of identification strengths (non, weakly, semi-strongly, and strongly identified). The proposed statistic is straightforward to calculate and has standard normal limiting distribution, which is *asymptotically pivotal*.

3.3.1 J -Test

As noted in Remark 3.2.10, Theorem 3.2.11 holds without any essential restrictions. The J -statistic is constructed based on the following theorem, which is closely related to Theorem 3.2.11.

⁴⁰There is a slight difference from the case here that Andrews and Mikusheva (2012) consider a “normalized” model and assume $\sigma_T = 1$ for any T in the ARMA(1,1) model therein.

Theorem 3.3.1 (*J*-statistic). *Given a drifting sequence $\{\xi_T\}_{T=1}^\infty \subset \Xi_{1,1}^*$, we have*

$$f_T(\theta_T) - \delta_T = T^{-1/2}(1 - \phi_T^2)C_T(-\phi_T) + O_p(T^{-1}).$$

As a corollary, let

$$J(\xi_T) \triangleq \frac{T^{1/2}(f_T(\theta_T) - \delta_T)}{(1 - \phi_T^2)^{1/2}}, \quad (3.114)$$

then

$$J(\xi_T) \xrightarrow{d} \mathcal{N}(0, 1).$$

*Thus, the *J*-statistic is asymptotically pivotal.*

By previous results, the above theorem follows immediately. Based on the *J*-statistic, we have the *J*-test which is straightforward to implement in practice. Given hypothesis $H_0 : \xi = \xi_T \triangleq (\delta_T, \theta_T)'$, significance level α , and observations $\{Y_{t,T}\}_{t=0}^T$, we calculate $f_T(\theta_T)$ as

$$\arg \min_{\delta} \tilde{F}_T(\delta, \theta | \{Y_{t,T}\}_{t=0}^T), \quad (3.115)$$

where \tilde{F}_T is defined in (3.29). After calculating the value of $J(\xi_T)$ using Formula (3.114), we do inference according to the standard normal distribution.

3.3.2 Simulation Results

For the purpose of comparison, we run simulations of *J*-tests and quasi-likelihood ratio (QLR) tests at the same time, and the simulation results are depicted as below. In Figures (3.2)–(3.19), $\xi^* = (\delta^*, \theta^*)' = (\frac{d^*}{\sqrt{T}}, \theta^*)'$ are the true parameter values, and $\xi_0 = (\delta_0, \theta_0)' = (\frac{d_0}{\sqrt{T}}, \theta_0)'$ are the parameter values in the null hypotheses. Significance level $\alpha = 0.05$. Coverage probability and power are calculated based on 2000 simulations. Each figure

presents simulation results for sample size $T = 100, 250$, and 500 .

According to Theorem 3.3.1, we see that the coverage probability of J -test should be very close to $1 - \alpha$, which is confirmed by Figures (3.2)–(3.8). From these figures, we see that the coverage probability of the J -test is very precise when neither ϕ^* nor θ^* is close to 1 in absolute value. In this case, the coverage probability of the QLR test is a little (about 1 percentage point) lower than $1 - \alpha$, i.e., the QLR test is slightly anticonservative in this case (if we use the usual critical value). However, when the absolute values of ϕ^* or θ^* get close to 1, the coverage probability of the J -test becomes lower than $1 - \alpha$, and that of the QLR test becomes very close to the nominal value $1 - \alpha$.

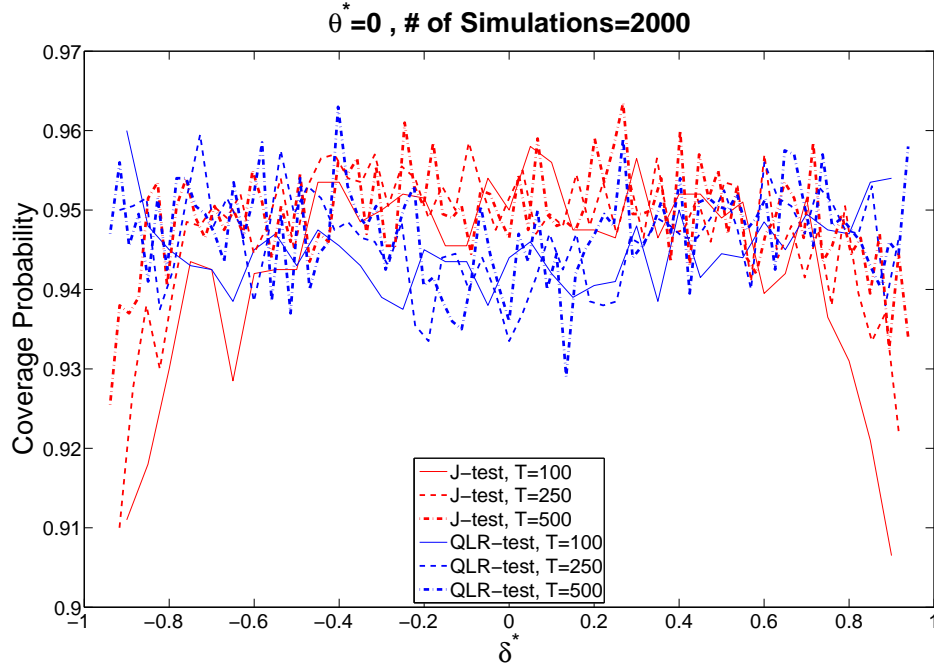


Figure 3.2: Coverage Probability of J - and QLR test, $\theta^* = 0$

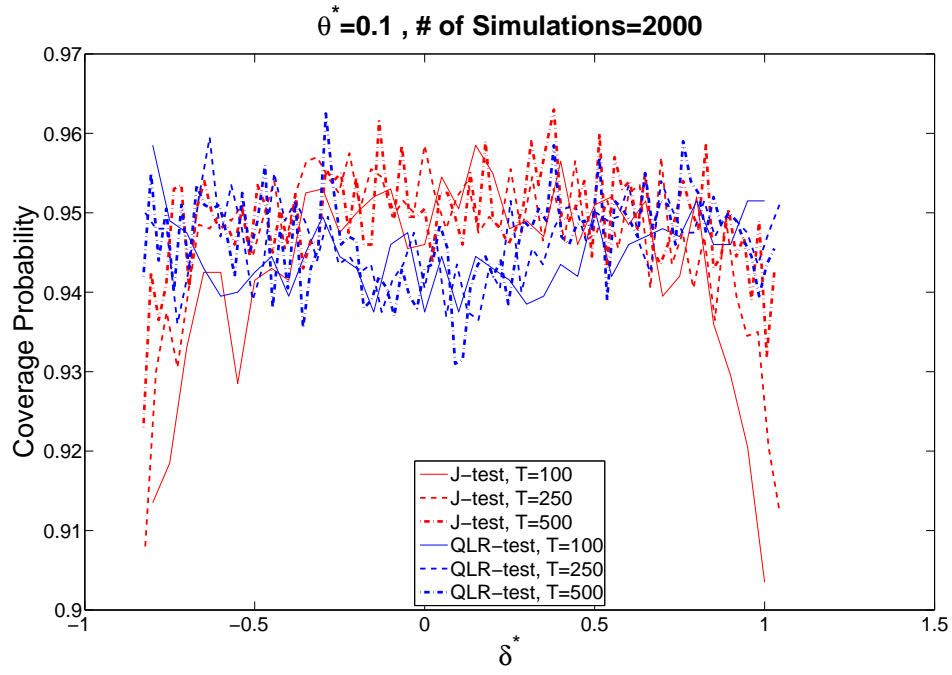


Figure 3.3: Coverage Probability of J - and QLR test, $\theta^* = 0.1$

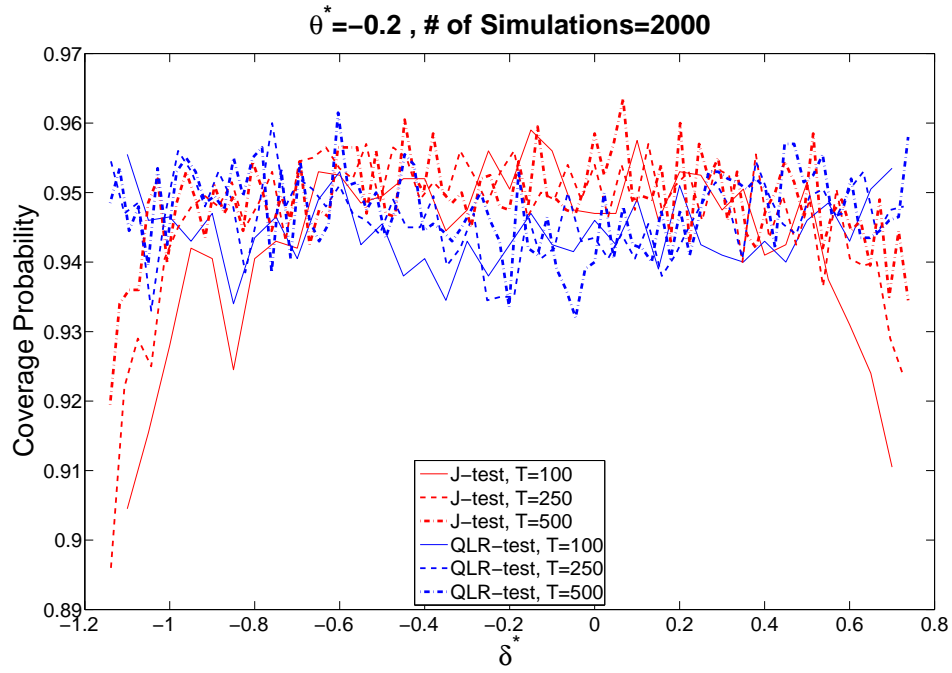


Figure 3.4: Coverage Probability of J - and QLR test, $\theta^* = -0.2$

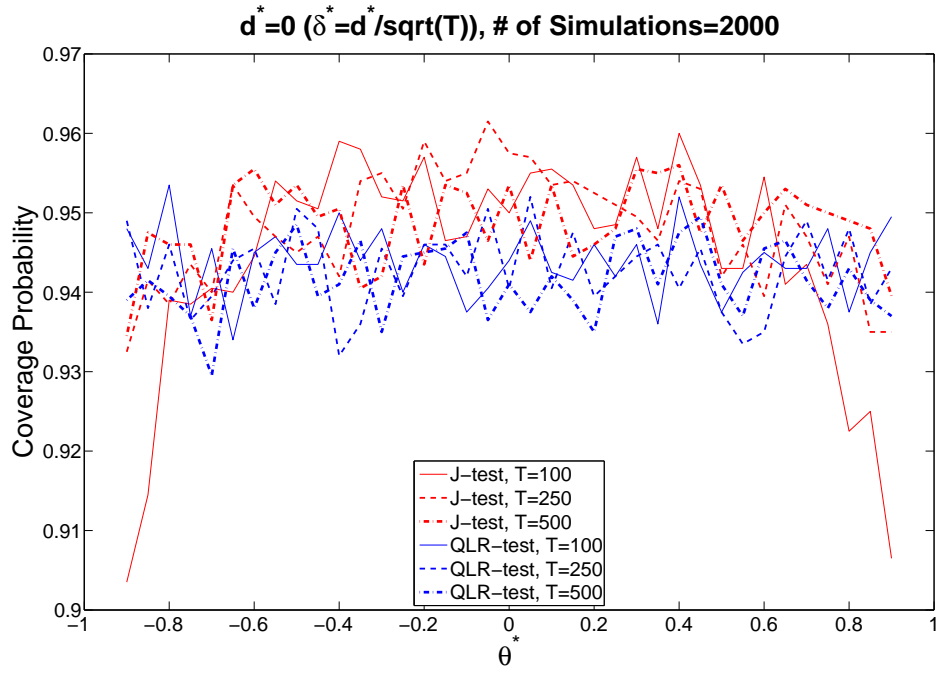


Figure 3.5: Coverage Probability of J - and QLR test, $d^* = 0$

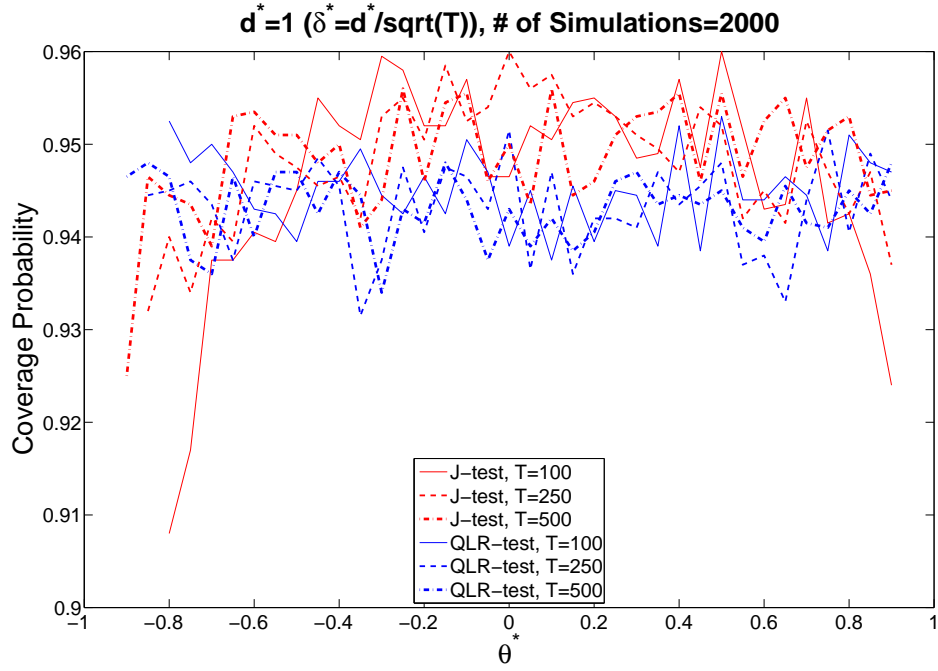


Figure 3.6: Coverage Probability of J - and QLR test, $d^* = 1$

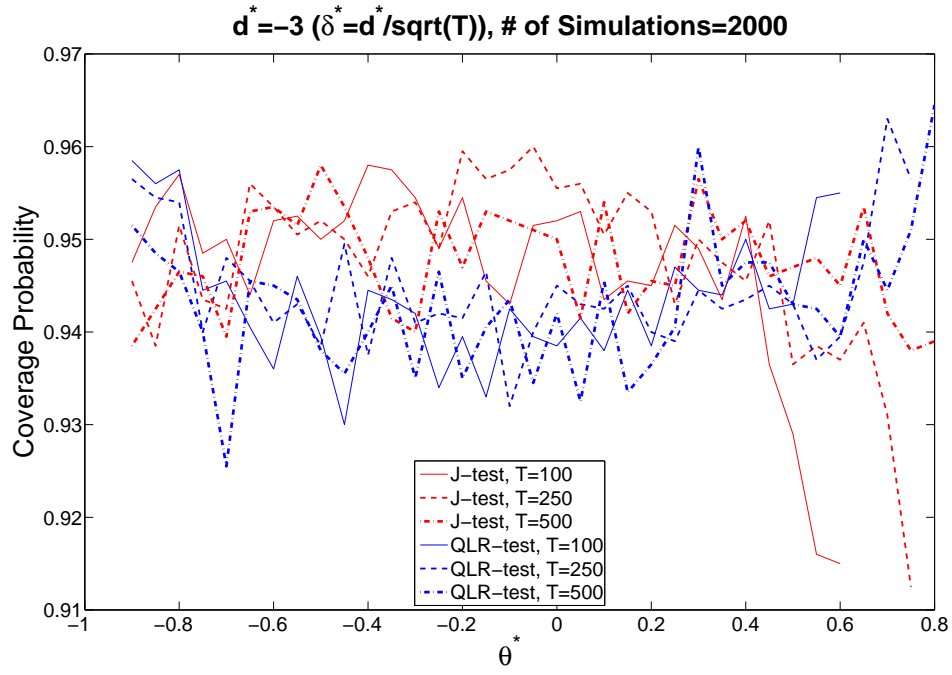


Figure 3.7: Coverage Probability of J - and QLR test, $d^* = -3$

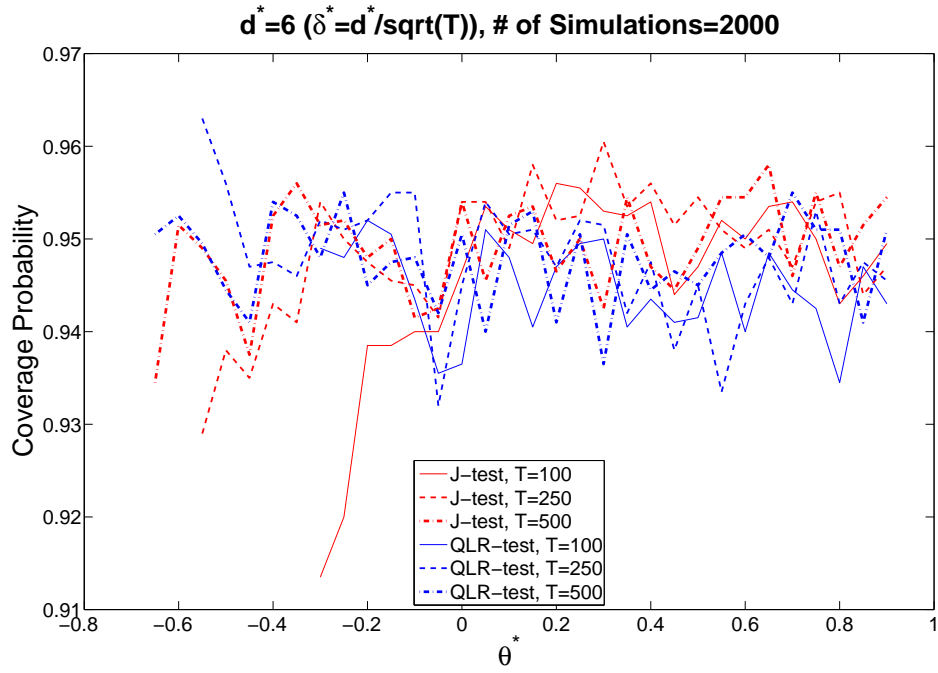


Figure 3.8: Coverage Probability of J - and QLR test, $d^* = 6$

As J -test mainly utilizes the strong identification property of δ^* , it is natural that the J -test behaves very well in the δ direction, and less well in the θ direction. Figures 3.9–3.13 confirms this result. For these figures, J -test performs better than the QLR test and provides more precise size and larger power. Figure 3.14 is essentially “the same” as Figure 3.5 (except for the graphs are “symmetric” about the horizontal line $y = 1/2$), as they are both characterizing the unidentified case.

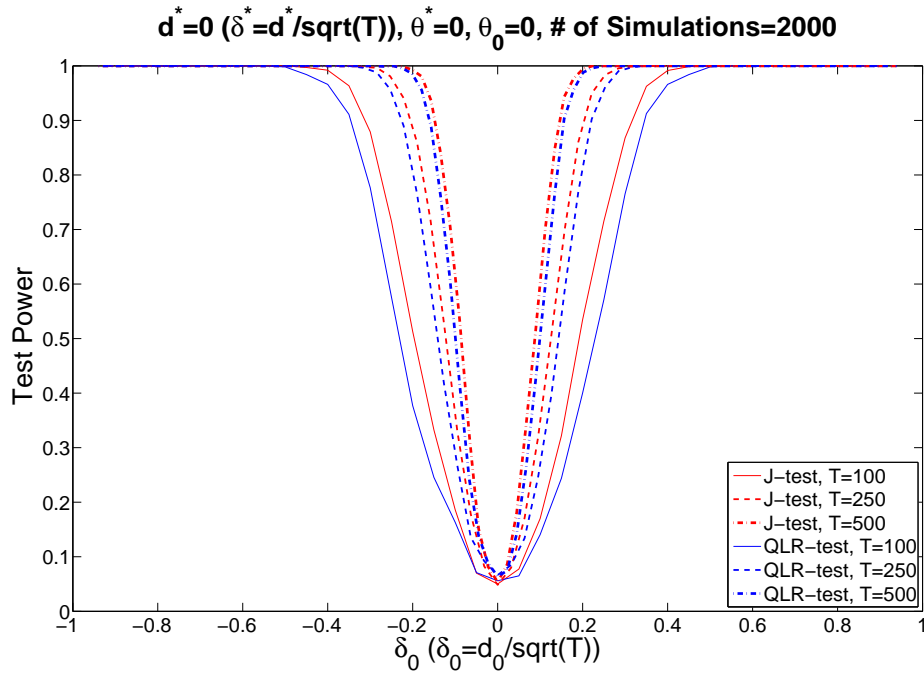


Figure 3.9: Power of J - and QLR test, $d^* = \theta^* = \theta_0 = 0$

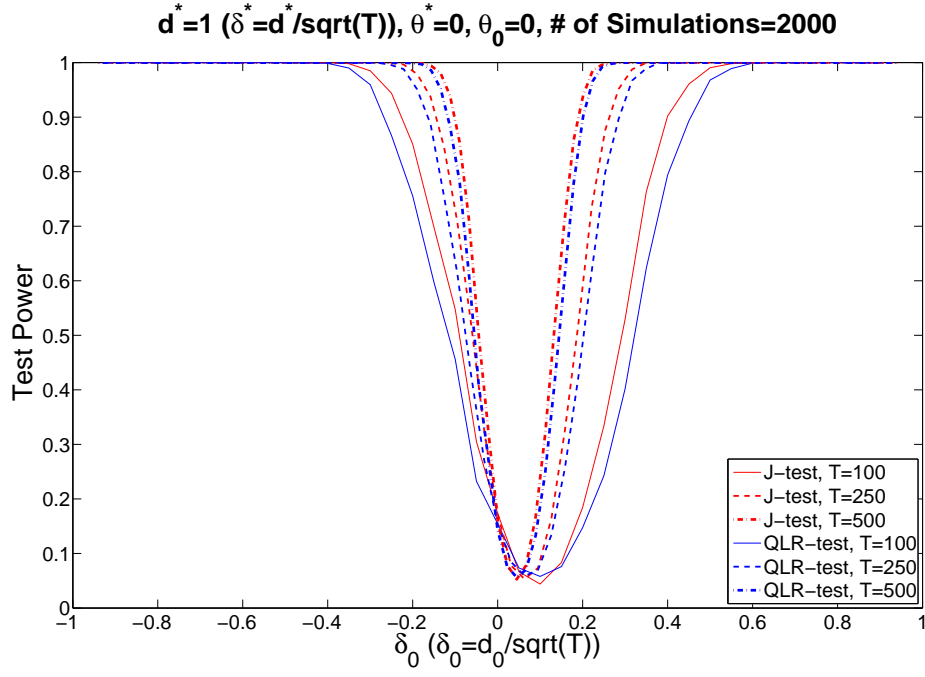


Figure 3.10: Power of J - and QLR test, $d^* = 1, \theta^* = \theta_0 = 0$

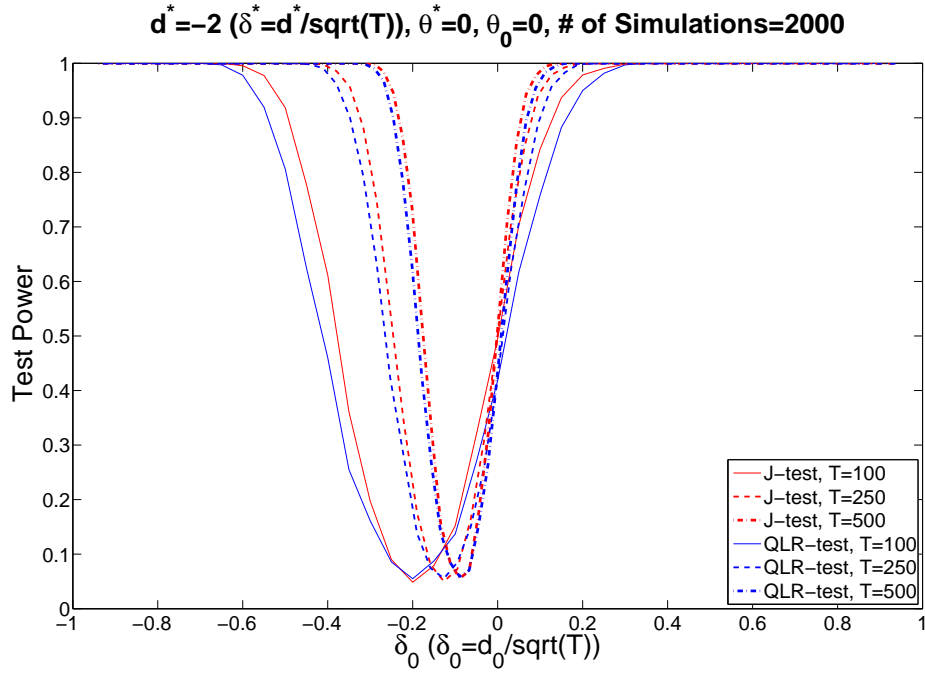


Figure 3.11: Power of J - and QLR test, $d^* = -2, \theta^* = \theta_0 = 0$

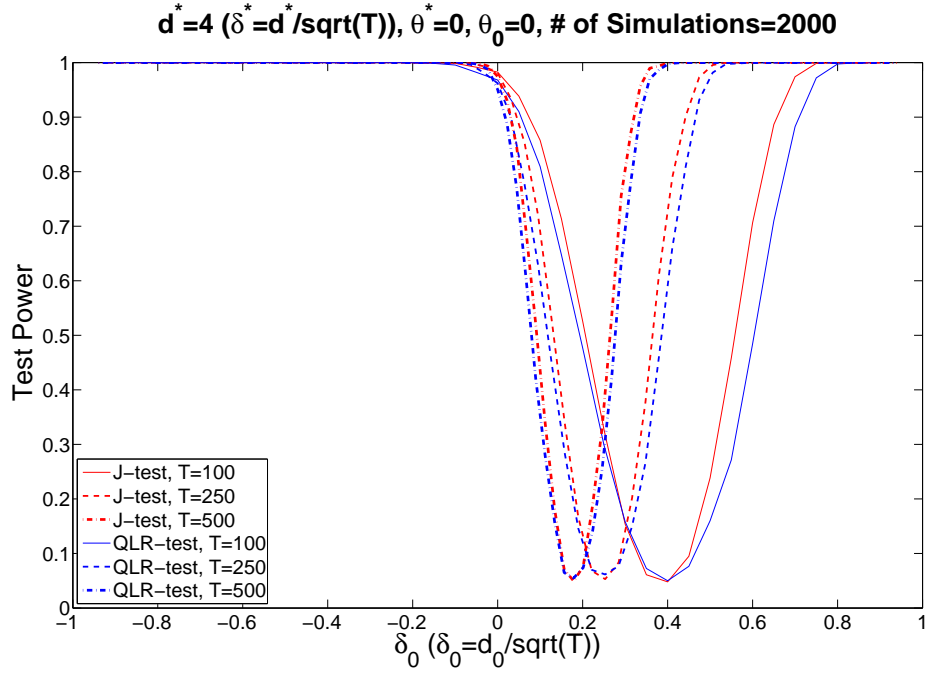


Figure 3.12: Power of J - and QLR test, $d^* = 4, \theta^* = \theta_0 = 0$

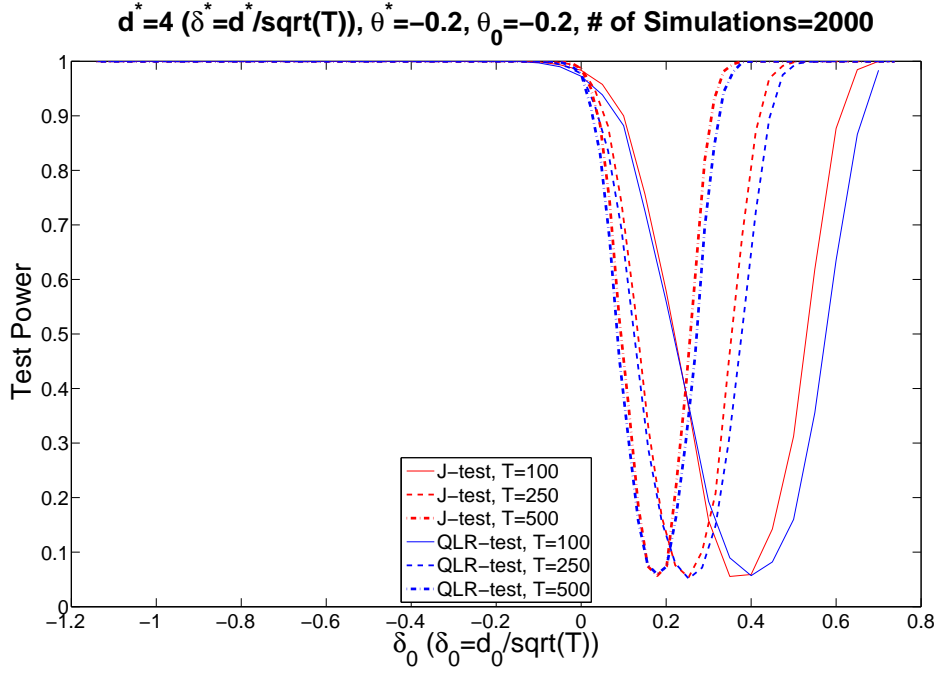


Figure 3.13: Power of J - and QLR test, $d^* = 4, \theta^* = \theta_0 = -0.2$

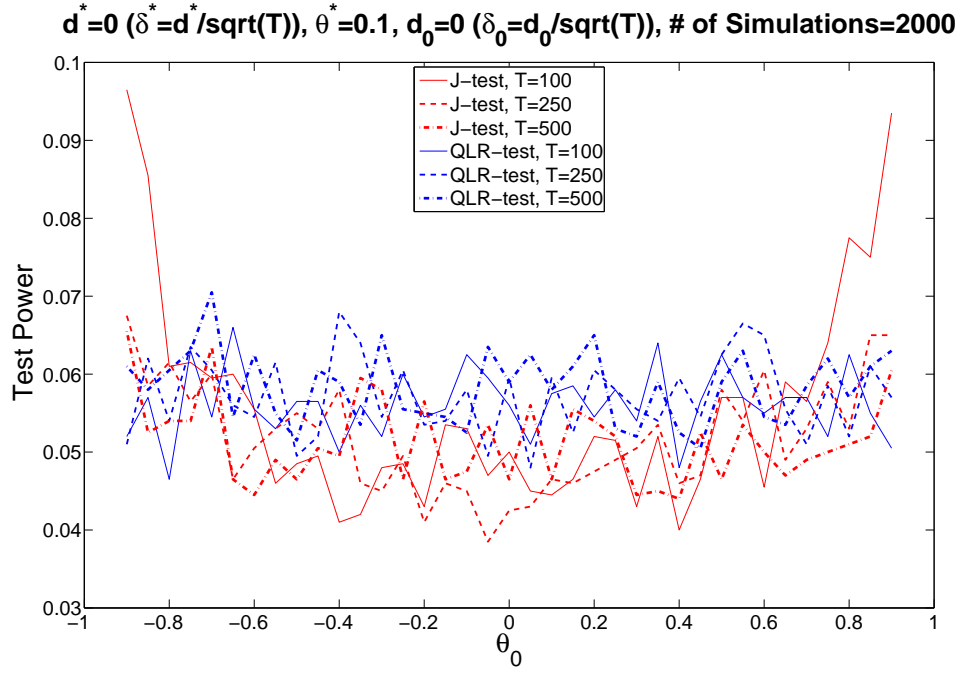


Figure 3.14: Power of J - and QLR test, $d^* = d_0 = 0, \theta^* = 0.1$

When the true parameter values are at or close to the root cancellation case, J -test performs very well, which can be seen from pervious figures and Figure 3.15, where the performances of J - and QLR test are very close even in the θ direction. When the true parameter values are not close to the root-cancelling case and θ^* and θ_0 are far apart, J -test performs less well than does the QLR test (see Figures (3.16)–(3.19)). From the simulations, we can see that unlike usual tests that performs better in strongly identified cases than in weakly and un- identified cases, the J -test performs particularly well for weakly and un-identified cases, which can be a very desirable property for some applications.

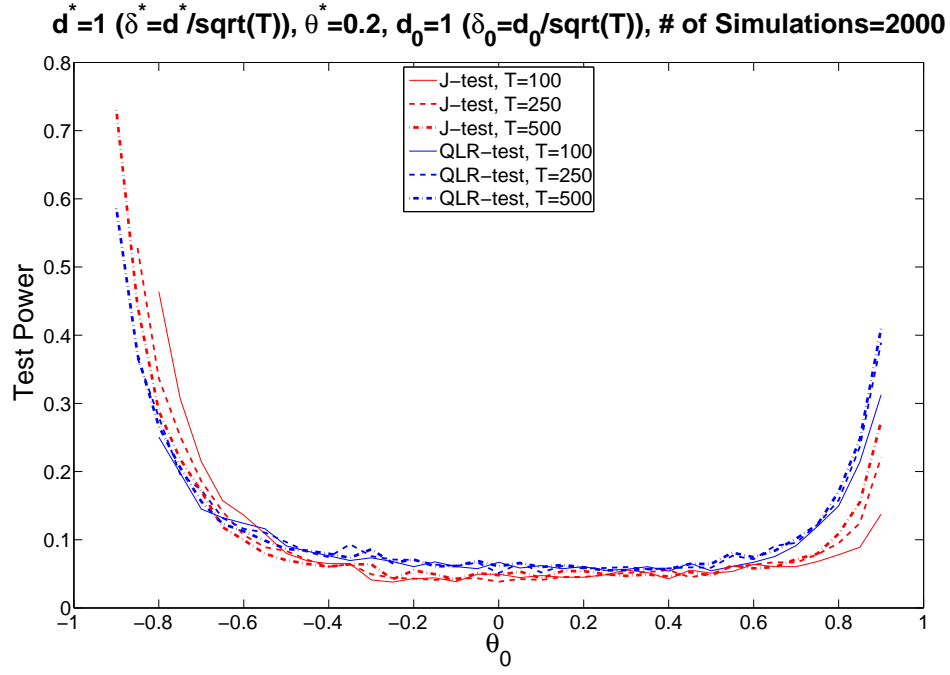


Figure 3.15: Power of J - and QLR test, $d^* = d_0 = 1, \theta^* = 0.2$

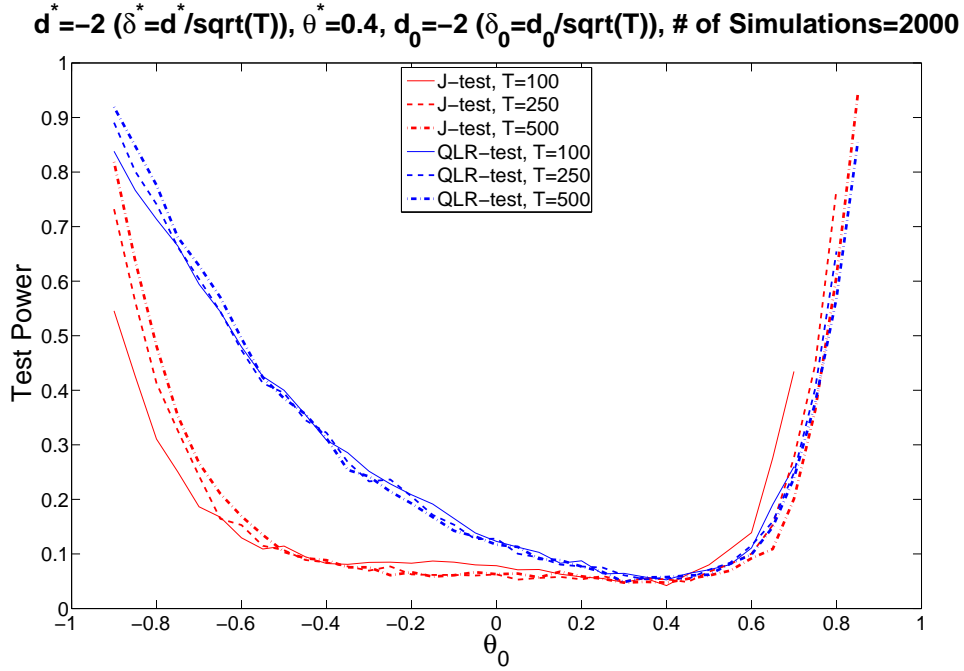


Figure 3.16: Power of J - and QLR test, $d^* = d_0 = -2, \theta^* = 0.4$

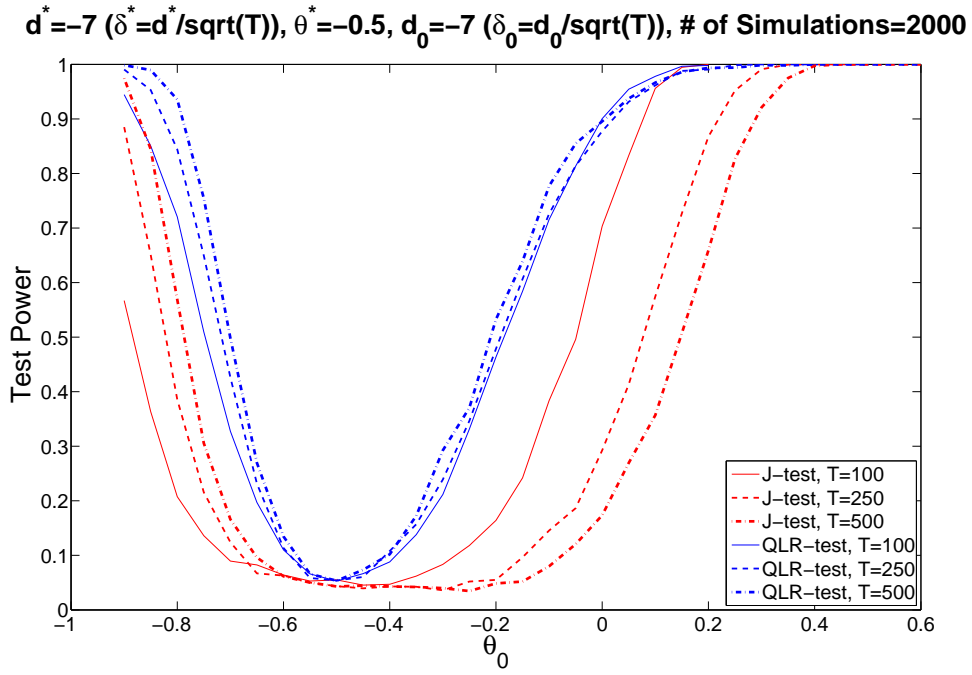


Figure 3.17: Power of J - and QLR test, $d^* = d_0 = -7, \theta^* = -0.5$

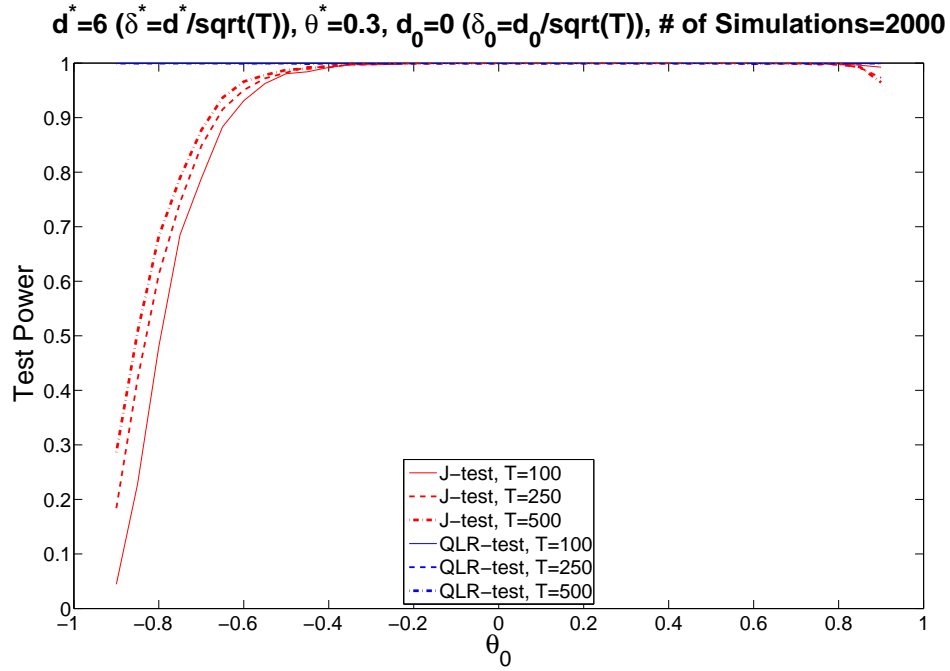


Figure 3.18: Power of J - and QLR test, $d^* = 6, d_0 = 0, \theta^* = 0.3$

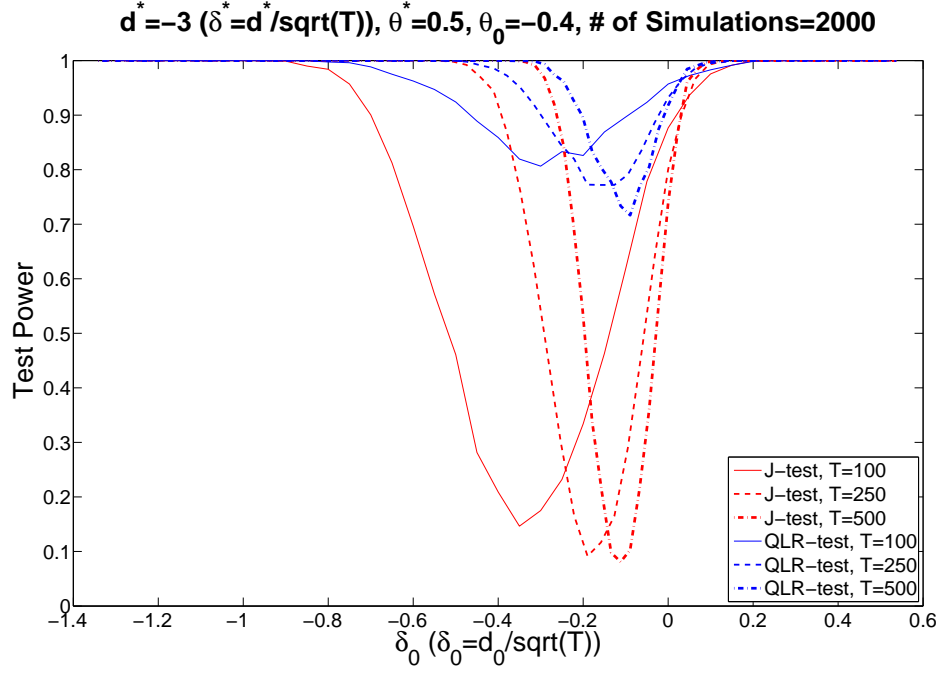


Figure 3.19: Power of J - and QLR test, $d^* = -3$, $\theta^* = 0.5$, $\theta_0 = -0.4$

3.4 Conclusion

This paper studies the estimation and testing of ARMA(1,1) models with root cancellation using the global approach. The idea of the global approach is quite intuitive. For the ARMA(1,1) case, minimizing the (quasi)log-likelihood function \mathcal{L}_T with respect to $(\xi, \sigma)'$ is converted to minimizing F_T over $\Xi_{1,1}^+$ first and then finding $\hat{\theta}_T$ to minimize \mathcal{L}_T . In the key step of minimizing F_T over $\Xi_{1,1}^+$, we minimize F_T in the direction of δ first, and the property of F_T guarantees that there is a unique minimum $\delta = f_T(\theta)$ for any given θ . Thus, we obtain a curve $c_T(\theta) \triangleq (f_T(\theta), \theta)'$ of the valleys of F_T . Then the minimum of F_T over $\Xi_{1,1}^+$ is the minimum of the function $\mathcal{V}_T(\theta) \triangleq F_T(c_T(\theta))$, which is pinned down by checking the behavior of $\mathcal{V}_T'(\theta)$ over Θ globally. In order to understand and simplify $\mathcal{V}_T'(\theta)$ over Θ ,

we explore the properties of a certain kind of (formal) power series in the lag operator L (called “ A -series”), the uniform convergence/boundedness (in probability) of a certain kind of stochastic series (called “ B -series”), and solve the first-order equations. For the no and weak identification cases, the asymptotic distribution is determined by the “shape analysis function” and an equation for $\hat{\delta}_T$. For the semi-strongly identified case, the asymptotic distribution is found by solving a stochastic linear equation which is derived from the strong transversal property of a deterministic part of the random function $\mathcal{V}'_T(\theta)$. The global approach introduced here can also be used to deal with the usual strongly identified (no root cancellation) case, if more terms are computed. (Those terms are higher order infinitesimals when $\lim_{T \rightarrow \infty} \delta_T = 0$, and are ignored in the analysis of the root cancellation case in this paper.)

Using the global approach, we prove Theorem 3.1.2 and 3.1.3, which give the necessary and sufficient conditions for a drifting sequence of true distributions to be non, weakly, semi-strongly, or strongly identified, or to have no asymptotic distribution. Thus, these theorem provide a complete classification of drifting sequences of true distributions in terms of asymptotic distributions (summarized in Table 3.1). Based on this classification, we show the equivalence classes of drifting sequences (Theorem 3.1.4). When the asymptotic distributions exist, the global approach gives the rates of convergence and asymptotic distributions. For the semi-strongly and strongly identified case, $\hat{\delta}_T - \delta_T = O_p(T^{-1/2})$ and $\hat{\theta}_T - \theta_T = O_p((T^{1/2}\delta_T)^{-1})$, and the asymptotic distributions $\pi_{\hat{\xi}_T}$ are bivariate normal. For the non and weakly identified case, $\hat{\delta}_T - \delta_T = O_p(T^{-1/2})$ and $\hat{\theta}_T - \theta_T = O_p(1)$, and the asymptotic distributions are not one of the commonly seen distributions. We see that the rate of convergence of $\hat{\delta}_T - \delta_T$ is always $T^{-1/2}$, but the rate of convergence of $\hat{\theta}_T - \theta_T$ is

$T^{-1/2}\delta_T^{-1}$. The global approach intuitively illustrates the mechanism how the strength of identification changes with $|\lim_{T \rightarrow \infty} T^{1/2}\delta_T|$. That is, based on the information obtained from the first-order equation for δ_T , the behavior of $\mathcal{V}'_T(\theta)$ over Θ shows how the strength of identification of θ_T intensifies from no identification to (semi-) strong identification as $|\lim_{T \rightarrow \infty} T^{1/2}\delta_T|$ increases from 0 to $+\infty$. Then the identification of θ_T together with the first-order equation for δ_T explains why the strength of identification of δ_T remains “unchanged” during the same process.

Based on the fact that the parameter δ_T is always strongly identified, we propose the J -statistic for conducting joint tests on the AR and MA parameters. The J -statistic is straightforward to calculate and has a standard normal limiting distribution, which is asymptotically pivotal, and the corresponding J -test is robust to the full range of identification strengths. The actual size of J -test is very close to the nominal value, and the test power is satisfactory in the δ direction. The performance of J -test is particularly good when the true parameter values are of or close to the root cancellation cases.

References

- ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica*, **55** (6), 1465–1471.
- (1992). Generic uniform convergence. *Econometric Theory*, **8** (2), 241–257.
- and CHENG, X. (2012a). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, **80** (5), 2153–2211.
- and — (2012b). Supplement to ‘estimation and inference with weak, semi-strong, and strong identification’. *Econometrica Supplemental Material*, **80**, http://www.econometricsociety.org/ecta/Supmat/9456_miscellaneous.pdf, http://www.econometricsociety.org/ecta/Supmat/9456_data_and_programs.zip.
- and PLOBERGER, W. (1996). Testing for serial correlation against an ARMA(1,1) process. *Journal of the American Statistical Association*, **91** (435), 1331–1342, Theory and Methods.
- ANDREWS, I. and MIKUSHEVA, A. (2012). Maximum likelihood inference in weakly identified DSGE models. Working Paper, MIT.
- ANSLEY, C. F. and NEWBOLD, P. (1980). Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics*, **13**, 159–183.
- BALKE, N. S. and GORDON, R. J. (1989). The estimation of prewar gross national product: Methodology and new evidence. *Journal of Political Economy*, **97**, 38–92.
- BANSAL, R., DITTMAR, R. F. and LUNDBLAD, C. T. (2005). Consumption, dividends, and the cross section of equity returns. *Journal of Finance*, **60** (4), 1639–1672.
- , KIKU, D. and YARON, A. (2010). Long run risks, the macroeconomy, and asset prices. *American Economic Review: Papers & Proceedings*, **100**, 542–546.
- , — and — (2012). An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, **1**, 183–221.
- and SHALIASTOVICH, I. (2012). *A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets*. 2012 Meeting Papers 778, Society for Economic Dynamics.
- and YARON, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, **59**, 1481–1509.

- BARRO, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, **121**, 823–866.
- (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review*, **99**, 243–264.
- and JIN, T. (2011a). On the size distribution of macroeconomic disasters. *Econometrica*, **79**, 1567–1589.
- and — (2011b). Supplement to ‘on the size distribution of macroeconomic disasters’. *Econometrica Supplemental Materials*, **79**, http://www.econometricsociety.org/ecta/Supmat/8827_data%20and%20programs.zip.
- and URSÚA, J. F. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity*, (Spring), 255–335.
- and URSÚA, J. F. (2010). Barro-Ursúa macroeconomic data. <http://scholar.harvard.edu/barro/publications/barro-ursua-macroeconomic-data>.
- and URSÚA, J. F. (2012). Rare macroeconomic disasters. *Annual Review of Economics*, **4**, 83–109.
- BEELER, J. and CAMPBELL, J. (2012). The long-run risks model and aggregate asset prices: An empirical assessment. *Critical Finance Review*, **1**, 141–182.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons, Inc., 2nd edn.
- BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series: Theory and Methods*. New York: Springer Science + Business Media, LLC, 2nd edn.
- BRUNNERMEIER, M. K. and NAGEL, S. (2008). Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals’ asset allocation. *American Economic Review*, **98**, 713–736.
- CAMPBELL, J. Y. (1993). Intertemporal asset pricing without consumption data. *American Economic Review*, **83**, 487–512.
- (2003). Consumption-based asset pricing. In G. Constantinides, M. Harris and R. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier, pp. 804–887.
- and COCHRANE, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, **107**, 205–251.
- and SHILLER, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, **1**, 195–228.
- CHAMPERNOWNE, D. G. (1953). A model of income distribution. *Economic Journal*, **63**, 318–351.
- CHEN, H. (2010). Macroeconomic conditions and the puzzles of credit spreads and capital structure. *The Journal of Finance*, **65**, 2171–2212.

- CHIAPPORI, P. A. and PAIELLA, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, **9** (6), 1021–1052.
- CHIB, S., NARDARI, F. and SHEPHARD, N. (2002). Markov chain monte carlo methods for stochastic volatility models. *The Journal of Econometrics*, **108**, 281–316.
- CHUNG, K. L. (2001). *A Course in Probability Theory*. Academic Press, 3rd edn.
- COGLEY, T. and STARTZ, R. (2012). Robust estimation of ARMA models with near root cancellation. Working Paper.
- COLACITO, R. and CROCE, M. M. (2011a). International risk-sharing with risk-sensitive rare events. Working Paper.
- and — (2011b). Risks for the long-run and the real exchange rate. *Journal of Political Economy*, **119**, 153–181.
- CONSTANTINIDES, G. M. (2008). Comment on ‘macroeconomic crises since 1870’ by R. J. Barro and J. F. Ursúa. *Brookings Papers on Economic Activity*, (Spring), 341–350.
- DONALDSON, J. B. and MEHRA, R. (2008). Risk-based explanations of the equity premium. In J. B. Donaldson and R. Mehra (eds.), *Handbook of the Equity Premium*, **2**, Amsterdam, The Netherlands: Elsevier.
- DUNSMUIR, W. and HANNAN, E. J. (1976). Vector linear time series models. *Adv. Appl. Prob.*, **8**, 339–364.
- DURRETT, R. (2010). *Probability: Theory and Examples*. Cambridge University Press, 4th edn.
- EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC.
- EPSTEIN, L. G. and ZIN, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, **57**, 937–969.
- FARHI, E., FRAIBERGER, S. P., GABAIX, X., RANCIERE, R. and VERDELHAN, A. (2009). Crash risk in currency markets. *NBER Working Paper*, (15062).
- and GABAIX, X. (2008). Rare disasters and exchange rates. *NBER Working Paper*, (13805).
- GABAIX, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, **1**, 255–293.
- (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, **127**, 645–700.
- , GOPIKRISHNAN, P., PLEROU, V. and STANLEY, H. E. (2003). A theory of power law distributions in financial market fluctuations. *Nature*, (423), 267–270.
- , —, — and — (2006). Institutional investors and stock market volatility. *Quarterly Journal of Economics*, **121**, 461–504.

- and IBRAGIMOV, R. (2011). Rank-1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, **29**, 24–39.
- and IOANNIDES, Y. (2004). The evolution of city size distributions. In V. Henderson and J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam, The Netherlands: North-Holland.
- and LANDIER, A. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics*, **123**, 49–100.
- GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd edn.
- GEWEKE, J. (2001). A note on some limitations of CRRA utility. *Economics Letters*, **71**, 341–345.
- GORODNICHENKO, Y., MIKUSHEVA, A. and NG, S. (2012). Estimators for persistent and possibly nonstationary data with classical properties. *Econometric Theory*, **28** (05), 1003–1036.
- GOURIO, F. (2008). Disasters and recoveries. *American Economic Review*, **98**, 68–73.
- (2012). Disaster risk and business cycles. *American Economic Review*, **102**, 2734–2766.
- GREENE, W. H. (2012). *Econometric Analysis*. Pearson Education, Inc., 7th edn.
- GRUBER, J. (2006). A tax-based estimate of the elasticity of intertemporal substitution. *NBER Working Paper*, (11945).
- GUVENEN, F. (2009). A parsimonious macroeconomic model for asset pricing. *Econometrica*, **77**, 1711–1750.
- HALL, R. E. (1988). Intertemporal substitution in consumption. *Journal of Political Economy*, **96**, 339–357.
- HANNAN, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Prob.*, **10**, 130–145.
- HANSEN, L. P., HEATON, J. C. and LI, N. (2008). Consumption strikes back? Measuring long-run risk. *Journal of Political Economy*, **116**, 260–302.
- HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**, 1163–1174.
- HODRICK, R. J. and PRESCOTT, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, **29**, 1–16.
- KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer Science+Business Media, Inc., 2nd edn.
- KLEIBERGEN, F. R. and HOEK, H. (2000). Bayesian analysis of ARMA models. *Tinbergen Institute Discussion Paper*, TI 2000-027/4.

- KOOP, G. and POTTER, S. M. (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, **74**, 763–789.
- LUCAS, R. E., JR. (1978). Asset prices in an exchange economy. *Econometrica*, **46**, 1429–1445.
- LUTTMER, E. G. J. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, **122**, 1103–1144.
- MA, J. and NELSON, C. R. (2008). Valid inference for a class of models where standard inference performs poorly; including nonlinear regression, ARMA, GARCH, and unobserved components. *Unpublished Manuscript*, Department of Economics, University of Washington.
- MALLOY, C. J., MOSKOWITZ, T. J. and VISSING-JORGENSEN, A. (2009). Long-run stockholder consumption risk and asset returns. *Journal of Finance*, **64**, 2427–2479.
- MANDELBROT, B. (1960). The Pareto-Levy law and the distribution of income. *International Economic Review*, **1**, 79–106.
- MEHRA, R. and PRESCOTT, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, **15**, 145–161.
- MIKUSHEVA, A. (2007). Uniform inference in autoregressive models. *Econometrica*, **75** (5), 1411–1452.
- MITZENMACHER, M. (2004a). A brief history of generative models for power law and log-normal distributions. *Internet Mathematics*, **1**, 226–251.
- (2004b). Dynamic models for file sizes and double Pareto distributions. *Internet Mathematics*, **1**, 305–333.
- NAKAMURA, E., SERGEYEV, D. and STEINSSON, J. (2012). Growth-rate and uncertainty shocks in consumption: Cross-country evidence. *NBER Working Paper*, (18128).
- , STEINSSON, J., BARRO, R. J. and URSÚA, J. (2013). Crises and recoveries in an empirical model of consumption disasters. *American Economic Journal: Macroeconomics*, **5**, 35–74.
- NELSON, C. R. and STARTZ, R. (2007). Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica*, **58**, 967–976.
- NEWBY, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, **59** (4), 1161–1167.
- PARETO, V. (1897). *Cours d'Economie Politique*, vol. 2. Paris, France: F. Pichou.
- PESARAN, M. H., PETTENUZZO, D. and TIMMERMAN, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, **73**, 1057–1084.
- POLLARD, D. (1990). *Empirical Processes: Theory and Applications*, CBMS Regional Conference Series in Probability and Statistics, vol. 2. Hayward, CA: Institute of Mathematical Statistics.

- PÖTSCHER, B. M. and PRUCHA, I. R. (1994). Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure. *Journal of Econometrics*, **60**, 23–63.
- QU, Z. (2011). Inference and specification testing in DSGE models with possible weak identification. Working Paper, Boston University.
- RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.
- REED, W. J. (2003). The Pareto law of incomes—an explanation and an extension. *Physica A*, (319), 469–486.
- RIETZ, T. A. (1988). The equity risk premium: A solution. *Journal of Monetary Economics*, **22**, 117–131.
- ROMER, C. D. (1986). Is the stabilization of the postwar economy a figment of the data. *American Economic Review*, **76**, 314–334.
- SEO, S. B. and WACHTER, J. A. (2013). Option prices in a model with stochastic disaster risk. Working Paper.
- STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65**, 557–586.
- STELLA, A. and STOCK, J. H. (2013). A state-dependent model for inflation forecasting. International Finance Discussion Papers, The Federal Reserve Board.
- STOCK, J. H. (1994). Unit roots, structural breaks and trends. In R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, 46, Elsevier B.V., pp. 2740–2843.
- and WATSON, M. W. (2003). Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002*, pp. 159–230.
- and WRIGHT, J. H. (2000). GMM with weak identification. *Econometrica*, **68**, 1055–1096.
- , — and YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, **20** (4), 518–529.
- TSAI, J. and WACHTER, J. A. (2012). Rare booms and disasters in a multi-sector endowment economy. Working Paper.
- URSÚA, J. F. (2010). Macroeconomic archaeology: Unearthing crises and long-term trends. Unpublished Manuscript, Harvard University.
- WACHTER, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, **68**, 987–1035.
- WEIL, P. (1990). Nonexpected utility in macroeconomics. *Quarterly Journal of Economics*, **105**, 29–42.
- WEITZMAN, M. L. (2007). Subjective expectations and asset-return puzzles. *American Economic Review*, **97**, 1102–1130.

Appendix A

Appendix to Chapter 1

A.1 Missing Data when $t = t_{i,0}$

When $t = t_{i,0}$, Formula (1.3) is not directly applicable, as $I_{i,t_{i,0}-1}$ is missing. Following the idea of (1.3), I calculate the following prior conditional probability instead

$$\begin{aligned} & \Pr(I_{i,t_{i,0}} = 1 | I_{W,t_{i,0}}) \\ = & \Pr(I_{i,t_{i,0}} = 1 | I_{i,t_{i,0}-1} = 0, I_{W,t_{i,0}}) \Pr(I_{i,t_{i,0}-1} = 0 | I_{W,t_{i,0}}) \\ & + \Pr(I_{i,t_{i,0}} = 1 | I_{i,t_{i,0}-1} = 1, I_{W,t_{i,0}}) \Pr(I_{i,t_{i,0}-1} = 1 | I_{W,t_{i,0}}) \\ = & p_{bW}^{I_{W,t_{i,0}}} p_{bI}^{1-I_{W,t_{i,0}}} \Pr(I_{i,t_{i,0}-1} = 0 | I_{W,t_{i,0}}) + (1 - p_e) \Pr(I_{i,t_{i,0}-1} = 1 | I_{W,t_{i,0}}). \end{aligned} \tag{A.1}$$

For simplicity, I further assume

$$\Pr(I_{i,t_{i,0}-1} = 1 | I_{W,t_{i,0}}) = \Pr(I_{i,t_{i,0}-1} = 1),$$

where the prior probability $\Pr(I_{i,t_{i,0}-1} = 1)$ is estimated by p_i , the fraction of event periods in all the periods studied for country i . So

$$\Pr(I_{i,t_{i,0}} = 1 | I_{W,t_{i,0}}) = p_{bW}^{I_{W,t_{i,0}}} p_{bI}^{1-I_{W,t_{i,0}}} (1 - p_i) + (1 - p_e) p_i, \tag{A.2}$$

and I impose the restriction that $p_i \in (0, 0.3)$.

A.2 Specification of Four Simulation Chains

In order to accurately estimate the model and assess convergence, I run four independent simulation chains in a way similar to that of NSBU. I specify two extreme scenarios: one is called “no-event scenario,” the other “all-event scenario.” For the no-event scenario, I set $I_{W,t} = 0$, $I_{i,t} = 0$, $x_{i,t,*} = y_{i,t,*}$, and $z_{i,t,*} = 0$ for i and t . For the all-event scenario, I set $I_{W,t} = 1$ and $I_{i,t} = 1$ for i and t , and extract a smooth trend using Hodrick-Prescott filter (Hodrick and Prescott, 1997, see) from the data. Let $y_{i,t,*}^{\tau}$ denote the trend component and $y_{i,t,*}^c$ the remainder, i.e.,

$$y_{i,t,*}^c = y_{i,t,*} - y_{i,t,*}^{\tau}.$$

I then let

$$z_{i,t,*} = \min(\max(-0.5, y_{i,t,*}^c), 0) \text{ and } x_{i,t,*} = y_{i,t,*} - z_{i,t,*}.$$

For each scenario, I specify two sets of initial values for parameters: one is called “lower values,” the other “upper values.” For the set of “lower values,” the initial parameter values are either close to their lower bounds or very low compared to their mean values. For the “upper values,” I have the opposite situation. Thus, the four sets of initial values of parameters for the four simulation chains are indeed far apart from each other.

A.3 An Auxiliary Treatment

The construction of the model determines that in the uneventful periods, the rare-event component (characterized by $\theta_{i,t,*}$ and $\phi_{i,t,*}$) will not pick up the long-run-risk effects, as

the event indicator $I_{i,t} = 0$. However, the long-run-risk component ($\triangleq u_{i,t,*} - \beta_i$) may pick up the rare-event effects if we do not give any additional treatment for it.

Before stating the additional treatment here, let me define several notations. Let

$$X \triangleq (X_1, \dots, X_{d_X})' = (p_W, p_{bW}, \dots, x_{1,t_{1,0},C}, x_{1,t_{1,0}+1,C}, \dots, x_{1,t_{1,0},GDP}, x_{1,t_{1,0}+1,GDP}, \dots)' \quad (\text{A.3})$$

be the random vector that contains all of the parameters and unobserved quantities as its components. The dimension d_X of X is in tens of thousands. Let

$$Y \triangleq (Y_1, \dots, Y_{d_Y})' = (y_{1,t_{1,0},C}, y_{1,t_{1,0}+1,C}, \dots, y_{1,t_{1,0},GDP}, y_{1,t_{1,0}+1,GDP}, \dots)' \quad (\text{A.4})$$

be the vector of observed variables for which I have data. The dimension d_Y of Y is equal to 10302 (= 4696 (for C) + 5606 (for GDP)). For country i , the vector of $v_{i,t,*}$'s is denoted by

$$V_{i,*} \triangleq (v_{i,t_{i,0},*}, \dots, v_{i,2009,*})'. \quad (\text{A.5})$$

In general, for a vector $Z \triangleq (Z_1, \dots, Z_{d_Z})$, let $Z_{[-Z_i]}$ denote the resulting vector when the component Z_i is removed from Z . Similarly, when vector W contains some component of Z , $Z_{[-W]}$ denotes the resulting vector when all the components contained in W are removed from Z . Then the *full conditional* posterior distribution of $v_{i,t,*}$, given all other parameters and unobserved quantities, is

$$p(v_{i,t,*} | X_{[-v_{i,t,*}]}, Y) \propto p(v_{i,t,*} | V_{i,*}[-v_{i,t,*}], \lambda_*, w_*) p(X_{[-(V_{i,*}, \lambda_*, w_*)]}, Y | V_{i,*}, \lambda_*, w_*). \quad (\text{A.6})$$

Clearly, the first term on the right hand side of Formula (A.6),

$$p(v_{i,t,*} | V_{i,*}[-v_{i,t,*}], \lambda_*, w_*),$$

is determined by the transition probability formula (1.8).

As the long-run-risk component fluctuates consistently, it is better to estimate it with only the data in the uneventful periods, as rare-event effects “contaminate” the data in event periods. To this end, I make the following assumption: When $I_{i,t} = 1$, i.e., when there is a rare event in country i at time t , it is assumed that

$$p(X_{[-(V_{i,*}, \lambda_*, w_*)]}, Y | V_{i,*}, \lambda_*, w_*)$$

doesn’t change with $v_{i,t,*}$, i.e.,

$$p(v_{i,t,*} | X_{[-v_{i,t,*}]}, Y) \propto p(v_{i,t,*} | V_{i,*}[-v_{i,t,*}], \lambda_*, w_*).$$

The above assumption is natural for the “disentanglement” of rare events and long-run risks. Without this assumption, it has to be the case that the variation of $\mu_{i,t,*}$ would pick up the event effects when a rare event occurs in country i . Note that it is only an auxiliary assumption. It makes the estimates more accurate, but doesn’t change them in any significant way.

A.4 Conversion to Monthly AR(1) Model

For AR(1) series $x_{t+1} = \rho x_t + \varepsilon_{t+1}$, $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$, the autocovariance function (ACF) of the *temporally aggregated series* $y_k = x_{(k-1)p+1} + \dots + x_{kp}$ is given by

$$\begin{aligned} \gamma_0 &= \frac{\sigma^2}{1-\rho^{2p}} \left[\sum_{i=1}^{p-1} 2\rho^{2p-i}(1+\rho+\dots+\rho^{i-1})(1+\rho+\dots+\rho^{p-i-1}) \right. \\ &\quad \left. + \sum_{i=1}^p \rho^{2(p-i)}(1+\rho+\dots+\rho^{i-1})^2 + \sum_{i=1}^{p-1} (1+\rho+\dots+\rho^{i-1})^2 \right], \\ \gamma_1 &= \rho^p \gamma_0 + \sigma^2 \sum_{i=1}^{p-1} \rho^{p-i}(1+\rho+\dots+\rho^{i-1})(1+\rho+\dots+\rho^{p-i-1}), \end{aligned}$$

and

$$\gamma_{j+1} = \rho^p \gamma_j$$

for $j \geq 1$. On the other hand, for any stationary series $\{z_k\}_k$, we can compute its autocovariance $\hat{\gamma}_k$. Given $m > 0$, let $\Gamma_m = (\gamma_0, \dots, \gamma_m)'$ and $\hat{\Gamma}_m = (\hat{\gamma}_0, \dots, \hat{\gamma}_m)'$, then the corresponding monthly autoregressive coefficient ρ and the standard deviation σ of innovations are given by

$$\arg \min_{(\tilde{\rho}, \tilde{\sigma})} \text{dist}(\Gamma_m(\tilde{\rho}, \tilde{\sigma}), \hat{\Gamma}_m).$$

In this study, I choose m to be 20 and “ $\text{dist}(\cdot, \cdot)$ ” the Euclidean metric. In Figure A.1, I depict the simulated ACF of the long-run growth rates estimated in the model and the ACF of the temporally aggregated series of the best fitting monthly model.

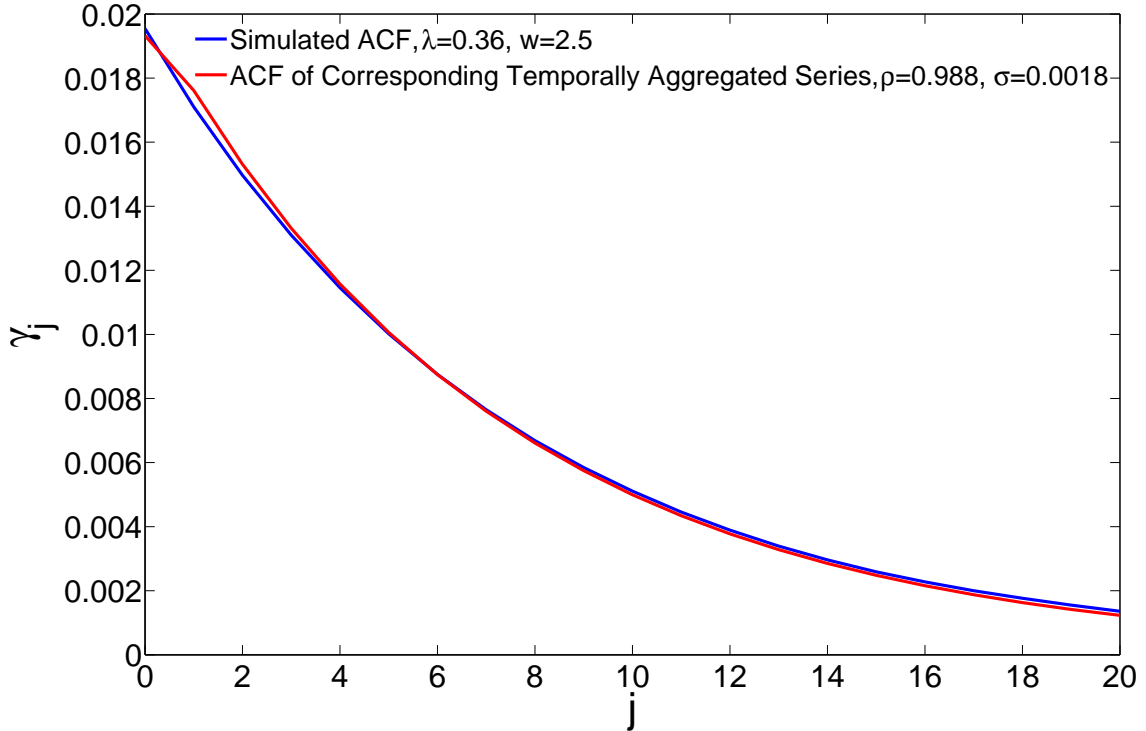


Figure A.1: Fitting the Autocovariance Function

A.5 Model Extension: Stochastic Volatility

The following empirical model has a long-run-risk part that is directly adapted from BY's setting.⁴¹ It contains an equation for log C (GDP)

$$y_{i,t} = x_{i,t} + z_{i,t} + \sigma_{\epsilon,i}\epsilon_{i,t}, \quad (\text{A.7})$$

an equation for potential level of log C (GDP)

$$\Delta x_{i,t} = \mu_i + \chi_{i,t-1} + \sigma_{i,t-1}\eta_{i,t} + I_{i,t}\theta_{i,t}, \quad (\text{A.8})$$

an equation for event gap

$$z_{i,t} = \rho_z z_{i,t-1} - I_{i,t}\theta_{i,t} + I_{i,t}\phi_{i,t} + \sigma_{v,i}v_{i,t}, \quad (\text{A.9})$$

an equation for persistent predictable component

$$\chi_{i,t+1} = \rho_\chi \chi_{i,t} + \varphi_e \sigma_{i,t} e_{i,t+1}, \quad (\text{A.10})$$

and an equation for the uncertainty risk

$$\sigma_{i,t+1}^2 = \sigma_i^2 + \rho_\sigma (\sigma_{i,t}^2 - \sigma_i^2) + \sigma_{\omega,i} \omega_{i,t+1}, \quad (\text{A.11})$$

where

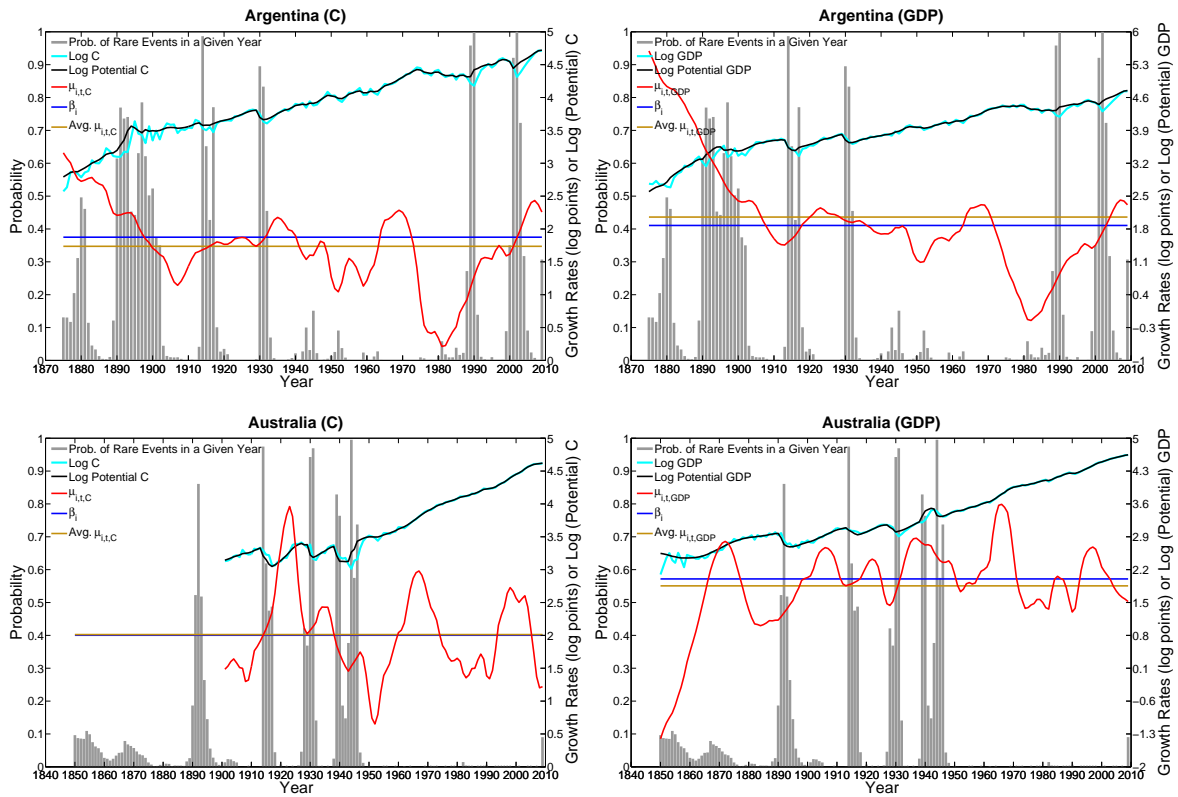
$$\epsilon_{i,t+1}, \eta_{i,t+1}, v_{i,t+1}, e_{i,t+1}, \omega_{i,t+1} \sim \text{i.i.d.} \mathcal{N}(0, 1).$$

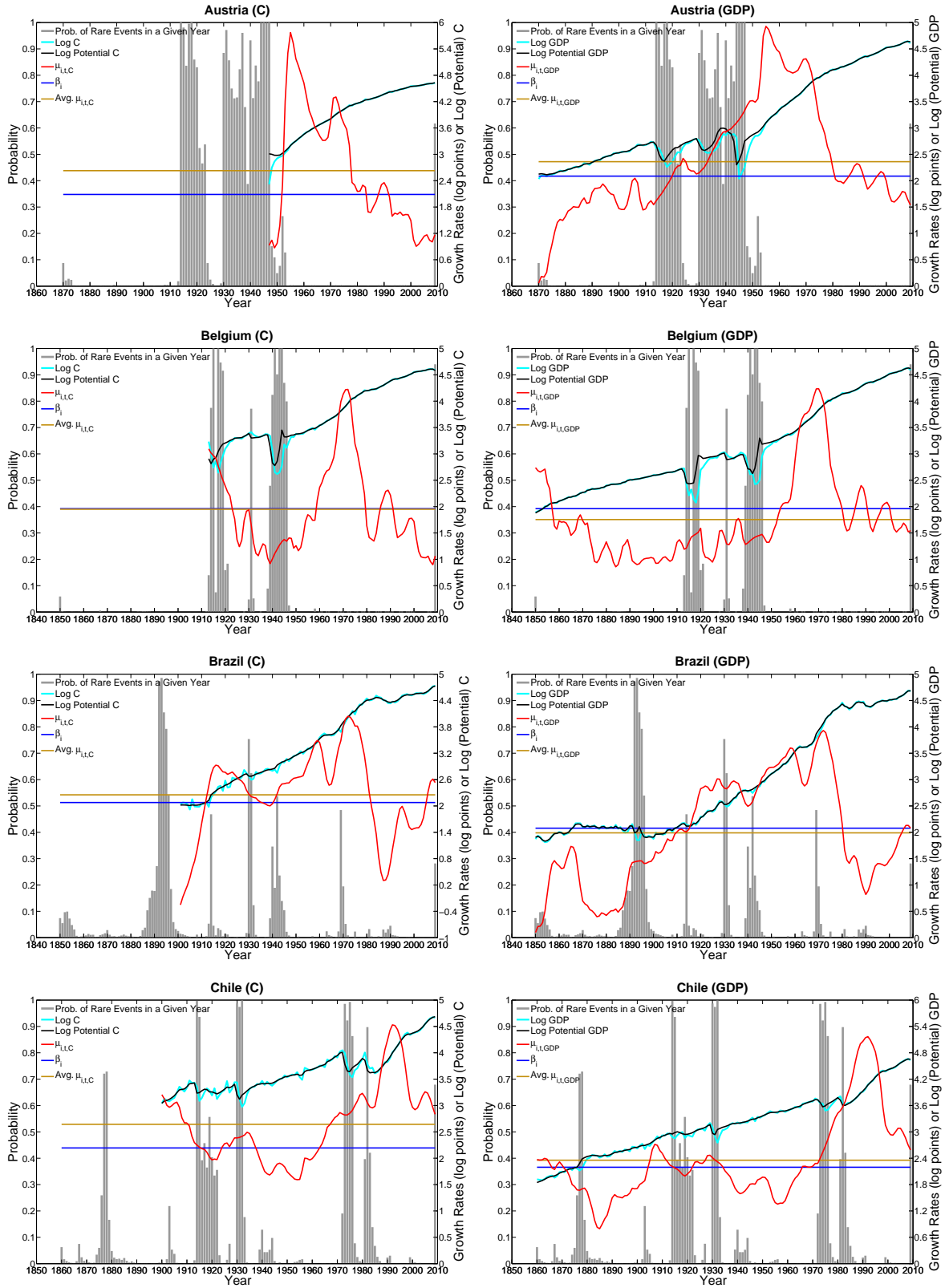
⁴¹Nakamura, Sergeyev and Steinsson (2012) study an empirical model of fluctuating growth rates and uncertainty shocks whose setting is based on BY. The main differences between the model here and the one in Nakamura, Sergeyev and Steinsson (2012) are: (1) their model does not have a rare-event part, while this one does; (2) their model has a global persistent predictable component in addition to country specific counterparts, while this one only has country specific components; (3) their model have two kinds of stochastic volatility processes—one for each country and one for the world, while this one only has the country specific stochastic volatility processes; (4) their model assume that the innovations to the stochastic volatility processes for all the countries have the same variance, while this one allow the variance to vary across countries.

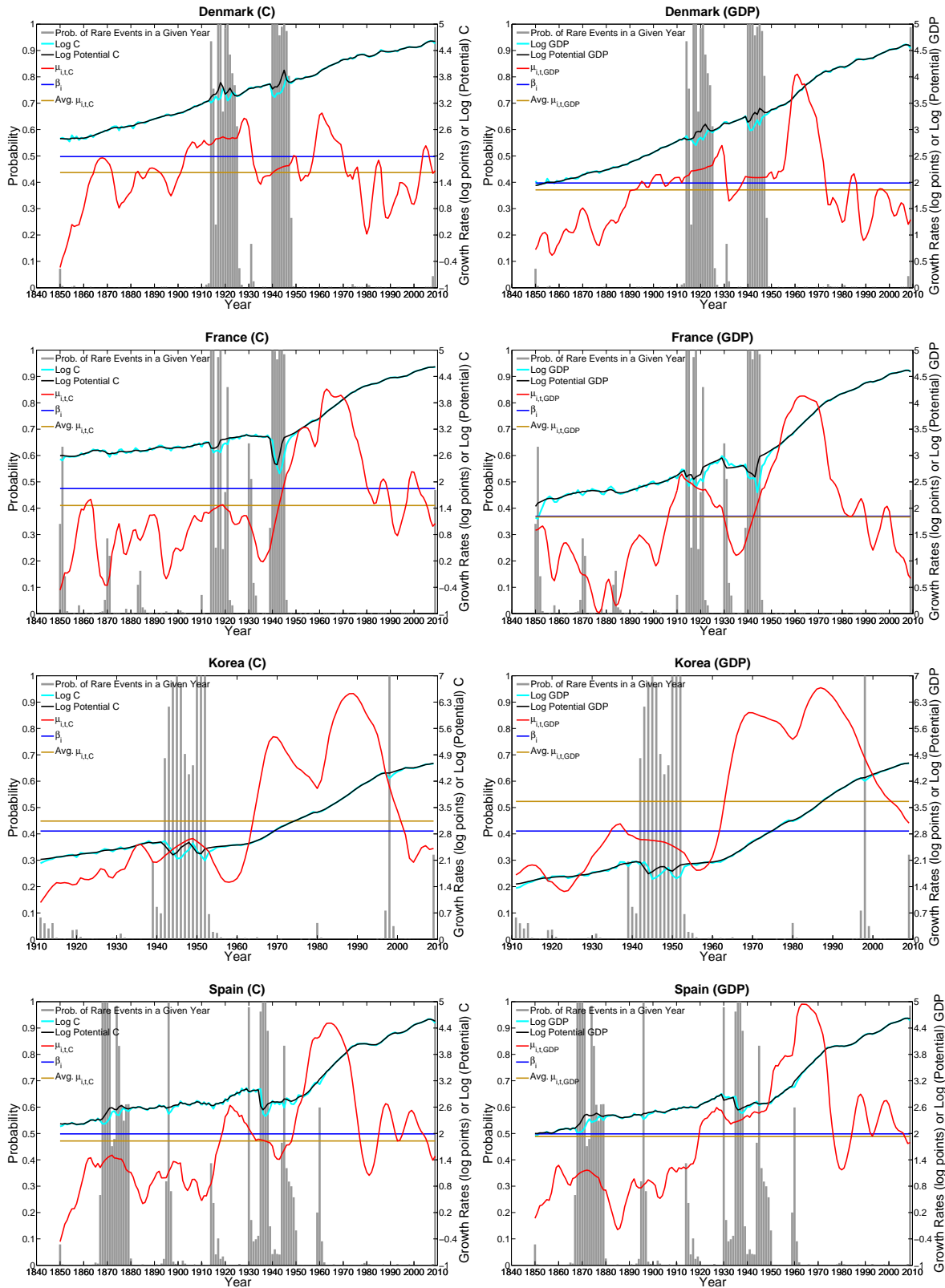
Compared to the model in this paper, the model listed above has more equations and parameters, so its identification and estimation can be more challenging.

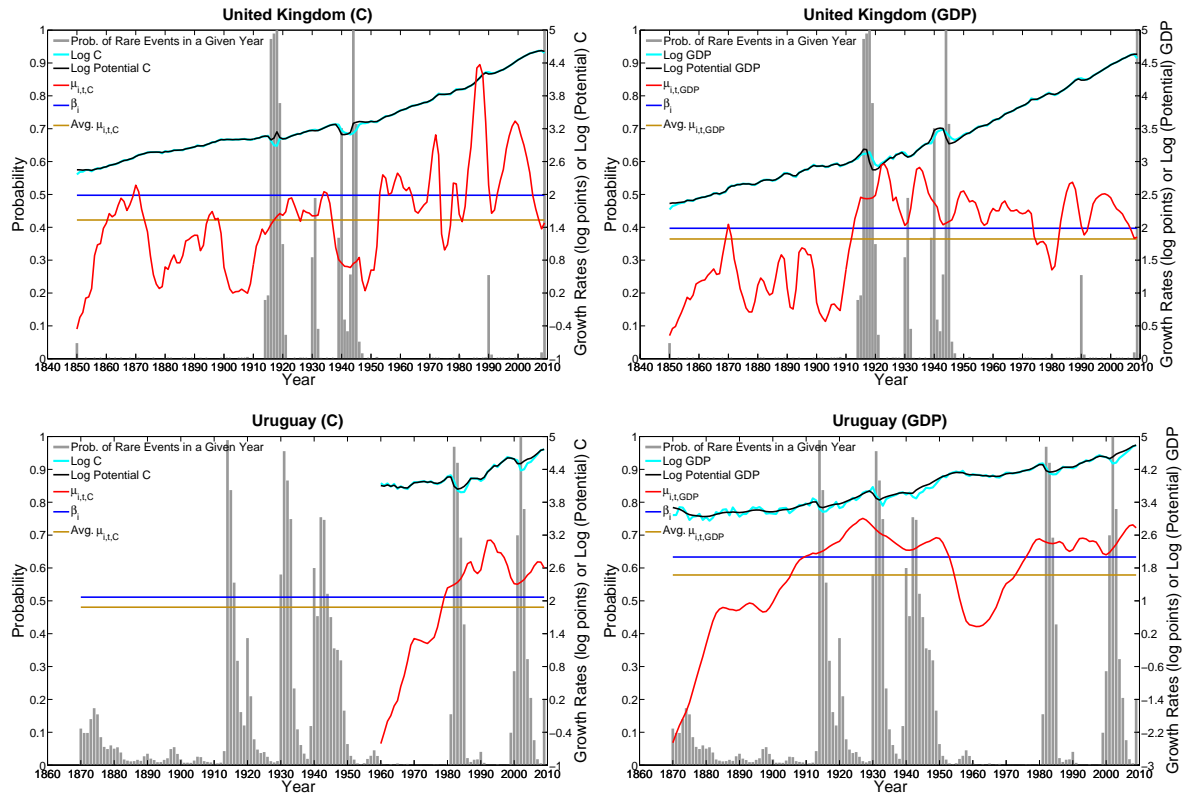
A.6 Supplemental Figures

1. Estimated probability of rare events, long-run growth rates, and potential consumption (output) for some economies explored in this study.

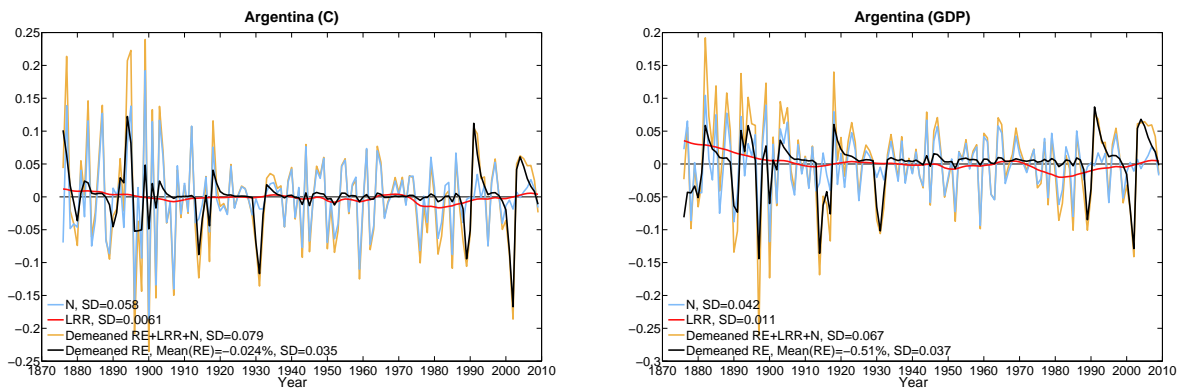


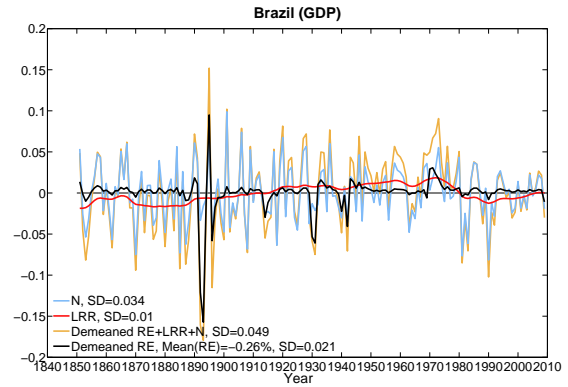
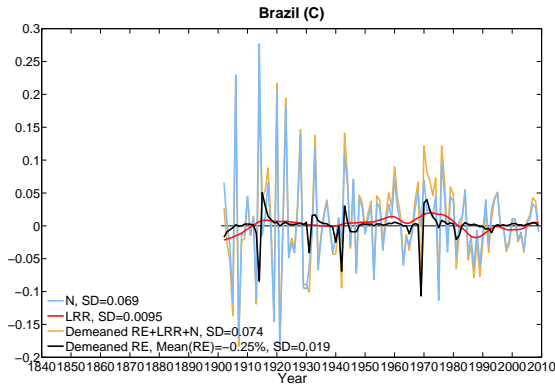
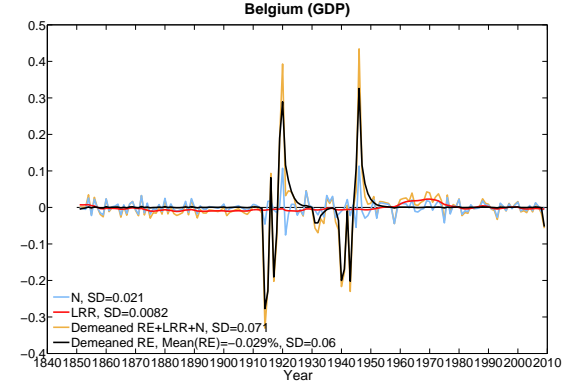
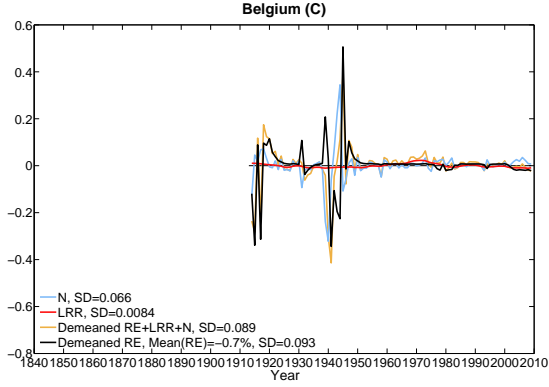
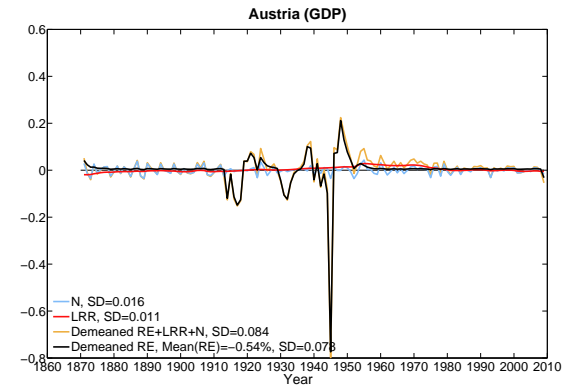
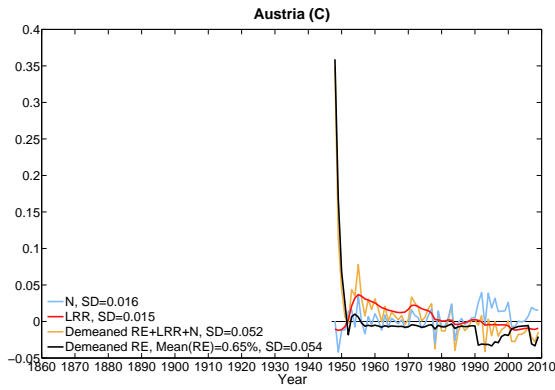
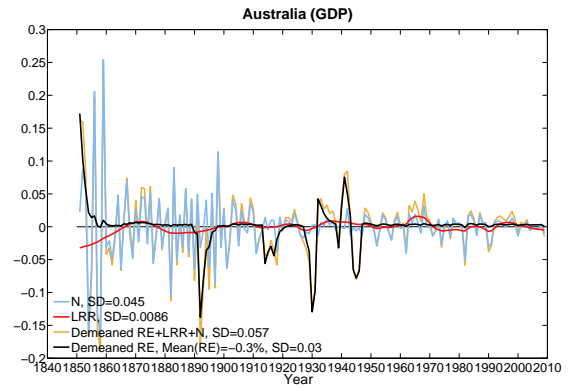
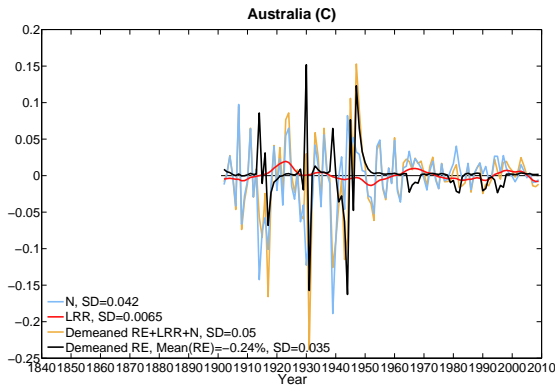


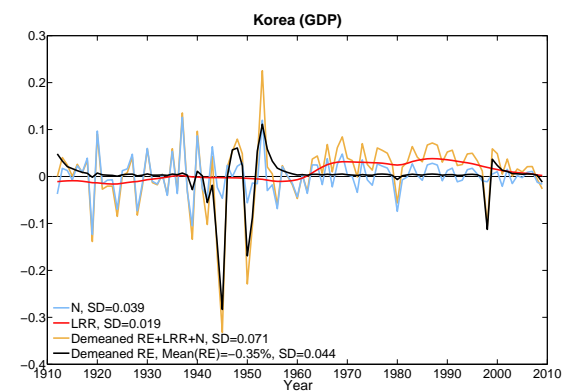
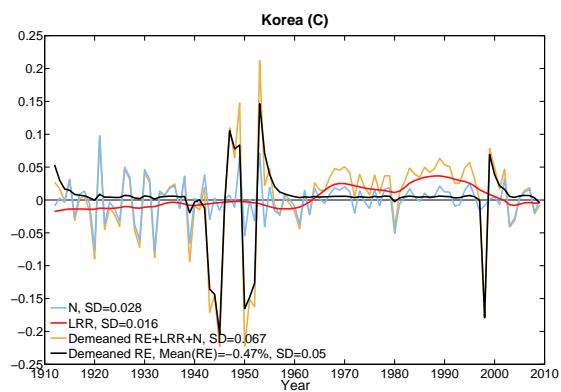
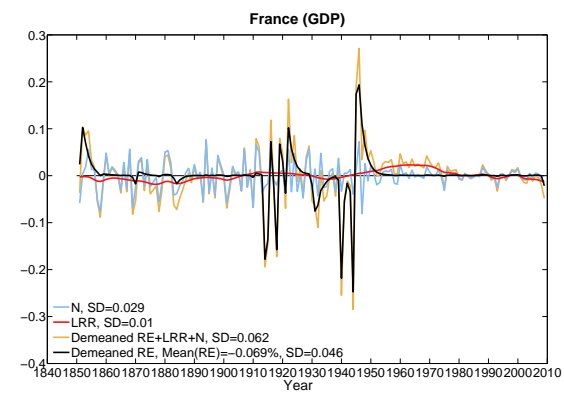
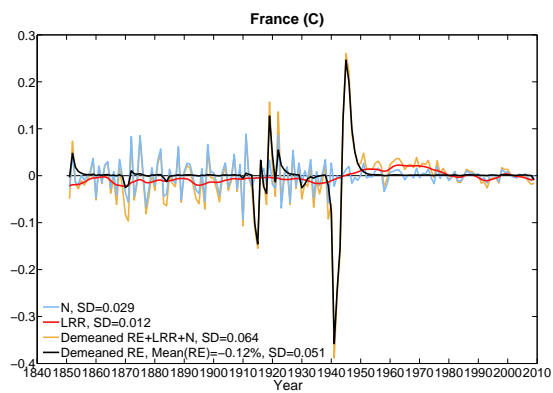
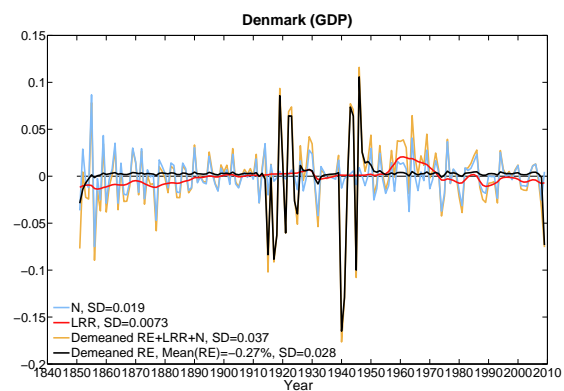
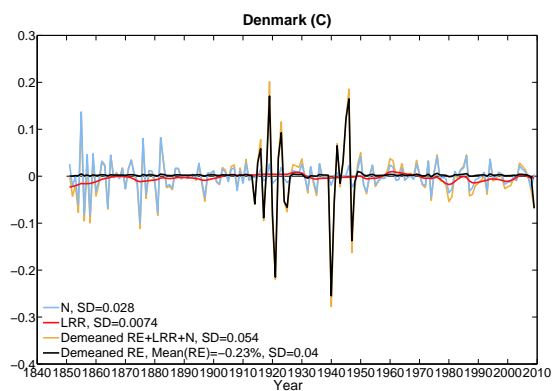
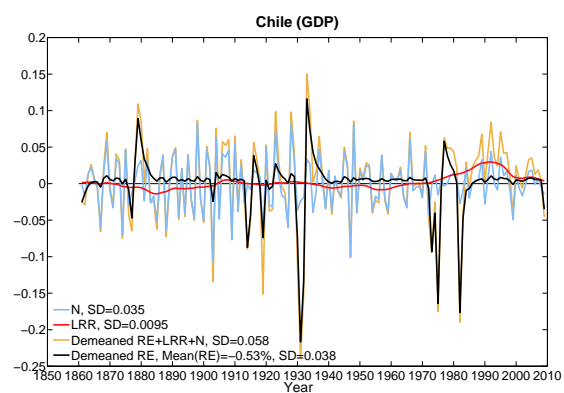
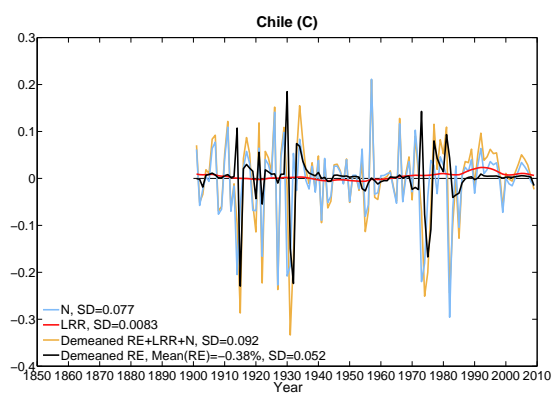


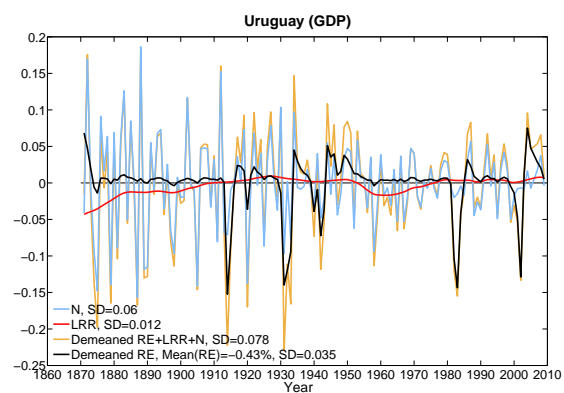
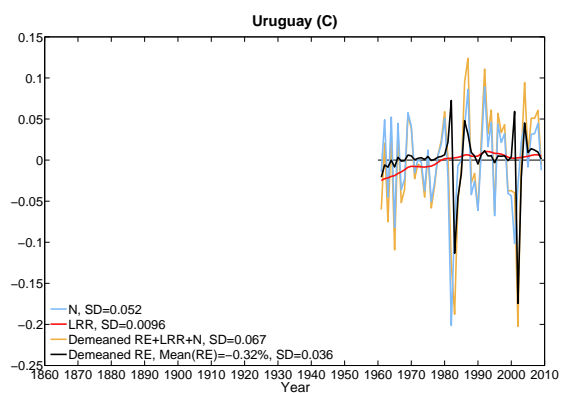
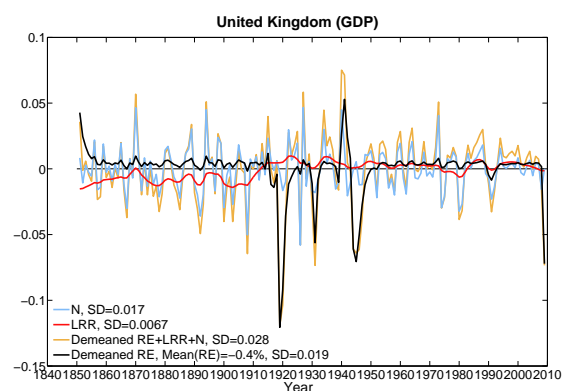
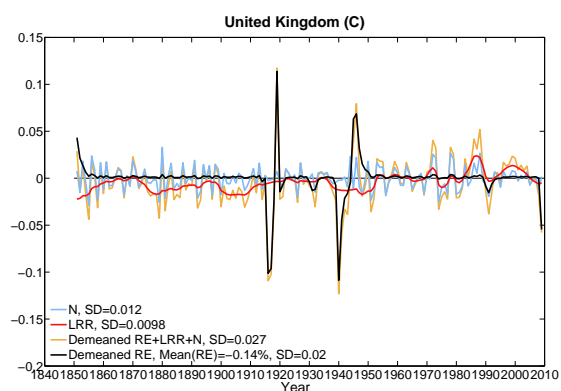
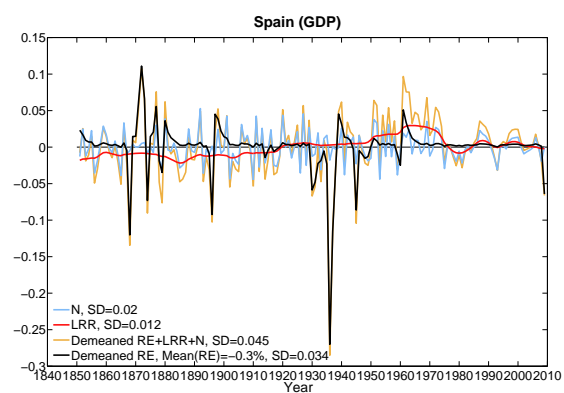
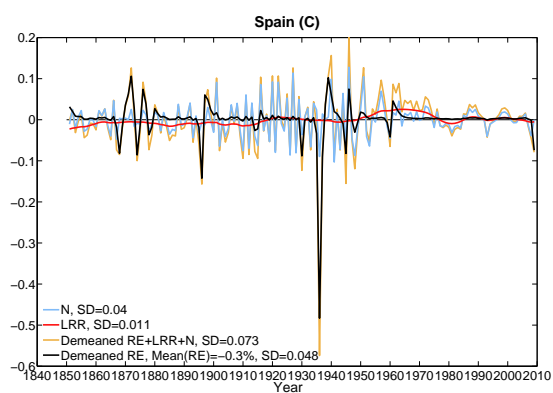


2. Decomposition of demeaned consumption (output) growth gaps for some economies explored in this study.









Appendix B

Appendix to Chapter 3

B.1 Propositions

Proposition B.1.1. *Let $\{X_T, T \in \mathbb{Z}_+\}$ and $\{Y_{T,i}, T, i \in \mathbb{Z}_+\}$ be random k -vectors such that*

1. $Y_{T,i} \xrightarrow{d} Y_i$ as $T \rightarrow \infty$ for every $i \in \mathbb{Z}_+$,
2. $Y_i \xrightarrow{d} Y$ as $i \rightarrow \infty$, and
3. $\lim_{i \rightarrow \infty} \limsup_{T \rightarrow \infty} \Pr(|X_T - Y_{T,i}| > \epsilon) = 0$ for every $\epsilon > 0$.

Then

$$X_T \xrightarrow{d} Y, \text{ as } T \rightarrow \infty. \quad (\text{B.1})$$

For a proof of the above proposition, see, for instance, Brockwell and Davis (2006, Chap. 6).

Proposition B.1.2. *R.v.'s $X_T \xrightarrow{a.s.} 0$ if and only if*

$$\Pr(|X_T| > \epsilon \text{ i.o.}) = 0 \text{ for any } \epsilon > 0. \quad (\text{B.2})$$

In the above proposition, “i.o.” stands for *infinitely often*. For a proof of it, see, for instance, Chung (2001, Chap. 4).

Proposition B.1.3 (Generic Uniform Convergence in Probability). *Given metric space (Θ, d_Θ) and sequence $\{H_T\}_{T \in \mathbb{Z}_+}$ as in (3.52). For the following properties:*

- (i) (Total Boundedness) (Θ, d_Θ) is a totally bounded metric space,
- (ii) (Pointwise Convergence in Probability) $H_T(\theta) = o_p(1)$ for any $\theta \in \Theta$,
- (iii) (Stochastic Equicontinuity) The sequence $\{H_T\}_{T \in \mathbb{Z}_+}$ is stochastically equicontinuous on Θ ,
- (iv) (Uniform Convergence in Probability) $H_T(\theta) = o_{p.u.}(1)$,

we have

1. (i), (ii), and (iii) \Rightarrow (iv),
2. (iv) \Rightarrow (ii) and (iii).

For a proof of the above proposition, see Andrews (1992). A similar version of the above result is proved in Newey (1991).

Proposition B.1.4. *Given $\beta^* = (\phi^{*'}, \theta^{*'})' = (\phi_1^*, \dots, \phi_p^*, \theta_1^*, \dots, \theta_q^*)' \in B_{p,q}^\circ$ ⁴² $\sigma^* > 0$, and an ARMA(p, q) process $Y_{\beta^*} = \{Y_{t,*}\}_{t \in \mathbb{Z}}$ with*

$$\Phi^*(z) = \prod_{k=1}^p (1 - \phi_k^* z) \text{ and } \Theta^*(z) = \prod_{k=1}^q (1 + \theta_k^* z) \quad (\text{B.4})$$

⁴²The definition of $B_{p,q}^\circ$ is a direct extension of that of $B_{1,1}^\circ$. That is,

$$B_{p,q}^\circ \triangleq \left\{ \beta = (\phi', \theta')' = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)' \in \mathbb{R}^{p+q} \mid |\phi_k|, |\theta_{k'}| \in (0, 1) \text{ for } k = 1, \dots, p, \text{ and } k' = 1, \dots, q, \{\phi_1, \dots, \phi_p\} \cap \{-\theta_1, \dots, -\theta_q\} = \emptyset \right\}. \quad (\text{B.3})$$

as the autoregressive and moving average polynomials, respectively. Let $Y_T = (Y_{1,*}, \dots, Y_{T,*})'$ be the vector of observations and Σ_T the covariance matrix of Y_T , i.e., $\Sigma_T = E(Y_T Y_T')$. Then the likelihood function of Y_T is

$$\tilde{L}_T(\Sigma_T | Y_T) = (2\pi)^{-T/2} \det(\Sigma_T)^{-1/2} \exp\left(-\frac{1}{2} Y_T' \Sigma_T^{-1} Y_T\right). \quad (\text{B.5})$$

Let $\tilde{\beta}_T$ be the MLE of β^* obtained by maximizing \tilde{L}_T over $B_{p,q}^\circ \times (0, \infty)$. Then

$$T^{1/2}(\tilde{\beta}_T - \beta^*) \xrightarrow{d} \mathcal{N}(0, W^{-1}(\beta^*)), \quad (\text{B.6})$$

where the matrix $W(\beta^*)$ can be calculated as follows. Let $\{U_t\}_t$ and $\{V_t\}_t$ be the autoregressive processes defined by

$$\Phi^*(L)U_t = \epsilon_t \text{ and } \Theta^*(L)V_t = \epsilon_t, \quad (\text{B.7})$$

where $\{\epsilon_t\}_t \sim i.i.d.(0,1)$. Let $Z_t = (U_t, U_{t-1}, \dots, U_{t-p+1}, V_t, V_{t-1}, \dots, V_{t-q+1})'$, then

$$W(\beta^*) = E(Z_t Z_t'). \quad (\text{B.8})$$

For a proof of the above proposition, see, for instance, Hannan (1973), Dunsmuir and Hannan (1976), and Brockwell and Davis (2006, Chap. 10).

Corollary B.1.5. *Given assumptions in Proposition B.1.4, for the ARMA(1,1) case, we have*

$$W(\beta^*) = \begin{pmatrix} \frac{1}{1-\phi^{*2}} & \frac{1}{1+\phi^*\theta^*} \\ \frac{1}{1+\phi^*\theta^*} & \frac{1}{1-\theta^{*2}} \end{pmatrix} \quad (\text{B.9})$$

and

$$\begin{aligned}
W^{-1}(\beta^*) &= \frac{1}{\det(W(\beta^*))} \begin{pmatrix} \frac{1}{1-\theta^{*2}} & \frac{-1}{1+\phi^*\theta^*} \\ \frac{-1}{1+\phi^*\theta^*} & \frac{1}{1-\phi^{*2}} \end{pmatrix} \\
&= \frac{1+\phi^*\theta^*}{\delta^{*2}} \begin{pmatrix} (1-\phi^{*2})(1+\phi^*\theta^*) & -(1-\phi^{*2})(1-\theta^{*2}) \\ -(1-\phi^{*2})(1-\theta^{*2}) & (1+\phi^*\theta^*)(1-\theta^{*2}) \end{pmatrix},
\end{aligned} \tag{B.10}$$

where $\det(W(\beta^*)) = \frac{1}{(1-\phi^{*2})(1-\theta^{*2})} - \frac{1}{(1+\phi^*\theta^*)^2}$. Let $\xi^* = (\delta^*, \theta^*)' = (\phi^* + \delta^*, \theta^*)'$, we have

$$T^{1/2}(\tilde{\xi}_T - \xi^*) \xrightarrow{d} \mathcal{N}(0, W(\xi^*)), \tag{B.11}$$

where

$$W(\xi^*) = \frac{1+\phi^*\theta^*}{\delta^{*2}} \begin{pmatrix} \delta^{*2}(1-\phi^*\theta^*) & \delta^*\phi^*(1-\theta^{*2}) \\ \delta^*\phi^*(1-\theta^{*2}) & (1+\phi^*\theta^*)(1-\theta^{*2}) \end{pmatrix}. \tag{B.12}$$

The above corollary immediately follows Proposition B.1.4.

B.2 Proofs

Proof of Lemma 3.2.1. One way to prove the result is to take the (formal) derivative of $A(r) = \sum_{j=0}^{\infty} r^j L^{j+1}$. Here we prove it by induction.

For $j = 0$, the result is trivially true. For $j = 1$, we have

$$A'(r) = -L(1-rL)^{-2}(-L) = L^2(1-rL)^{-2} = 1!A^2(r). \tag{B.13}$$

Namely, the result is true for $j = 1$. Assume that the result is true for $j = k$. Then for $j = k + 1$, we have

$$A^{(k+1)}(r) = (A^{(k)})'(r) = (k!A^{k+1})'(r) = (k+1)!A^k(r)A'(r) = (k+1)!A^{k+2}(r). \tag{B.14}$$

Thus, the result is proved for any $j \in \mathbb{N}$. \square

Proof of Lemma 3.2.2. Since

$$A(r')A(r) = \left(\sum_{i=0}^{\infty} r'^i L^{i+1} \right) \left(\sum_{j=0}^{\infty} r^j L^{j+1} \right) = L^2 \sum_{k=0}^{\infty} \left(\sum_{i=0}^k r'^i r^{k-i} \right) L^k, \quad (\text{B.15})$$

we have

$$\begin{aligned} (r' - r)A(r')A(r) &= L^2 \sum_{k=0}^{\infty} (r' - r) \left(\sum_{i=0}^k r'^i r^{k-i} \right) L^k = \sum_{k=0}^{\infty} (r'^{k+1} - r^{k+1}) L^{k+2} \\ &= \sum_{k=0}^{\infty} r'^{k+1} L^{k+2} - \sum_{k=0}^{\infty} r^{k+1} L^{k+2} = L[r'A(r') - rA(r)]. \end{aligned} \quad (\text{B.16})$$

Thus, the result is proved. \square

Proof of Lemma 3.2.3. (1) First, we consider $T^{-1/2} \sum_{t=1}^T [\epsilon_t(Q\epsilon_t) - \sigma^2 q_0]$. Let

$$Y_{T,i} = T^{-1/2} \sum_{t=1}^T \left[\epsilon_t \left(\sum_{j=0}^i q_j \epsilon_{t-j} \right) - \sigma^2 q_0 \right]. \quad (\text{B.17})$$

By the CLT for strictly stationary m -dependent sequences, we have that

$$Y_{T,i} \xrightarrow{d} Y_i \sim \mathcal{N}(0, (\sigma_4 - \sigma^4)q_0^2 + \sigma^4 \sum_{j=1}^i q_j^2), \text{ as } T \rightarrow \infty. \quad (\text{B.18})$$

Then by Proposition B.1.1 and Chebyshev's inequality, we have

$$T^{-1/2} \sum_{t=1}^T [\epsilon_t(Q\epsilon_t) - \sigma^2 q_0] \xrightarrow{d} \mathcal{N} \left(0, (\sigma_4 - \sigma^4)q_0^2 + \sigma^4 \sum_{j=1}^{\infty} q_j^2 \right), \text{ as } T \rightarrow \infty. \quad (\text{B.19})$$

Similarly, we have

$$T^{-1/2} \sum_{t=1}^T [\epsilon_t(R\epsilon_t) - \sigma^2 r_0] \xrightarrow{d} \mathcal{N} \left(0, (\sigma_4 - \sigma^4)r_0^2 + \sigma^4 \sum_{j=1}^{\infty} r_j^2 \right), \text{ as } T \rightarrow \infty. \quad (\text{B.20})$$

By Cramér-Wold device, we can show that the asymptotic distribution of

$$T^{-1/2} \left(\sum_{t=1}^T [\epsilon_t(Q\epsilon_t) - \sigma^2 q_0], \sum_{t=1}^T [\epsilon_t(R\epsilon_t) - \sigma^2 r_0] \right)'$$

is bivariate normal, and the result is proved.

(2) As Result (3.49) is a simple special case of (3.48), we consider it first. Let

$$Z_{T,i} = T^{-1} \sum_{t=1}^T \epsilon_t \left(\sum_{j=0}^i r_j \epsilon_{t-j} \right), \quad (\text{B.21})$$

then by the *strong law of large numbers (SLLN) for strictly stationary m -dependent sequences*, we know that $\lim_{T \rightarrow \infty} Z_{T,i} \xrightarrow{\text{a.s.}} \sigma^2 r_0$ for any $i \in \mathbb{Z}_T$. Using result in part (1) and Proposition B.1.2, we can show that $T^{-1} \sum_{t=1}^T \epsilon_t \left(\sum_{j=i+1}^{\infty} r_j \epsilon_{t-j} \right) \xrightarrow{\text{a.s.}} 0$ as $T \rightarrow \infty$, and the result of a.s. convergence in (3.49) follows.

However, there are tedious technical details in the above proof, and it only deals with a.s. convergence. Another proof, which is straightforward and leads to a.s. and L^1 convergence in (3.48) at the same time, is to use *Birkhoff's ergodic theorem* (see, for example, Durrett (2010, Chap. 7)).

(3) We need to show that $|v| < \infty$, and the rest of the proof is similar to that of part (1).

For any $k \in \mathbb{N}$,

$$\begin{aligned} \gamma(k) &= \text{Cov}(\tilde{W}_t, \tilde{W}_{t-k}) \\ &= (\sigma_4 - \sigma^4) \sum_{j=0}^{\infty} q_{j+k} r_{j+k} q_j r_j + \sigma^4 \sum_{j \neq j' \in \mathbb{N}} q_{j+k} r_{j'+k} q_j r_{j'} \\ &= (\sigma_4 - 2\sigma^4) \sum_{j=0}^{\infty} q_{j+k} r_{j+k} q_j r_j + \sigma^4 \sum_{j, j' \in \mathbb{N}} q_{j+k} r_{j'+k} q_j r_{j'}. \end{aligned} \quad (\text{B.22})$$

Thus,

$$\sum_{k=1}^{\infty} \gamma(k) = (\sigma_4 - 2\sigma^4) \sum_{j=0}^{\infty} \left(\sum_{k=1}^{\infty} q_{j+k} r_{j+k} \right) q_j r_j + \sigma^4 \sum_{j, j' \in \mathbb{N}} \left(\sum_{k=1}^{\infty} q_{j+k} r_{j'+k} \right) q_j r_{j'}. \quad (\text{B.23})$$

Let $\sum_{k=0}^{\infty} q_k^2 = M_q$, $\sum_{k=0}^{\infty} r_k^2 = M_r$, then $\sum_{k=0}^{\infty} |q_{j+k} r_{j'+k}| \leq (M_q + M_r)/2$ for any $j, j' \in \mathbb{N}$. So

we have

$$\begin{aligned}
& |\sum_{k=1}^{\infty} \gamma(k)| \\
& \leq |\sigma_4 - 2\sigma^4| \sum_{j=0}^{\infty} (\sum_{k=1}^{\infty} |q_{j+k} r_{j+k}|) |q_j r_j| + \sigma^4 \sum_{j,j' \in \mathbb{N}} (\sum_{k=1}^{\infty} |q_{j+k} r_{j'+k}|) |q_j r_{j'}| \\
& \leq |\sigma_4 - 2\sigma^4| (M_q + M_r)^2 / 4 + \sigma^4 [(M_q + M_r) / 2] (\sum_{j \in \mathbb{N}} |q_j|) (\sum_{j \in \mathbb{N}} |r_j|) \\
& < \infty.
\end{aligned} \tag{B.24}$$

Obviously, $\gamma(0) < \infty$, so $|v| < \infty$. \square

Proof of Theorem 3.2.5. The proof of this theorem is similar to that of Theorem 1 in Andrews (1992). \square

Proof of Lemma 3.2.6. Let $H_T(\theta) = T^{-1/2} \sum_{t=1}^T \tilde{W}_t(\theta)$ for each $T \in \mathbb{Z}$. By Theorem 3.2.5, the result is proved if conditions (i), (ii), and (iii) are satisfied. Condition (i) is obviously true as Θ is compact here. Condition (ii) holds according to Lemma 3.2.3. So we only need to show that the sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is stochastically equibounded. In fact, we will show a stronger result, i.e., that the sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is stochastically equicontinuous. We have

$$\begin{aligned}
& \tilde{W}_t(\theta) \\
& = (\sum_{j=0}^{\infty} q_j(\theta) r_j(\theta) \epsilon_{t-j}^2 - \sigma^2 \sum_{j=0}^{\infty} q_j(\theta) r_j(\theta)) + \sum_{j \neq j' \in \mathbb{N}} q_j(\theta) r_{j'}(\theta) \epsilon_{t-j} \epsilon_{t-j'} \\
& = \sum_{j,j' \in \mathbb{N}} s_{t,j,j'}(\theta)
\end{aligned} \tag{B.25}$$

for any $t \in \mathbb{Z}$, where

$$s_{t,j,j'}(\theta) = \begin{cases} q_j(\theta) r_j(\theta) (\epsilon_{t-j}^2 - \sigma^2), & \text{if } j = j'; \\ q_j(\theta) r_{j'}(\theta) \epsilon_{t-j} \epsilon_{t-j'}, & \text{otherwise.} \end{cases} \tag{B.26}$$

Then by *mean-value theorem*, for $\theta \neq \theta' \in \Theta$,

$$\tilde{W}_t(\theta') - \tilde{W}_t(\theta) = (\theta' - \theta) \sum_{j,j' \in \mathbb{N}} s'_{t,j,j'}(\vartheta_{\theta,\theta',t,j,j'}),$$

where $\vartheta_{\theta,\theta',t,j,j'} \in (\min\{\theta, \theta'\}, \max\{\theta, \theta'\})$, and

$$s'_{t,j,j'}(\vartheta) = \begin{cases} [q'_j(\vartheta)r_j(\vartheta) + q_j(\vartheta)r'_j(\vartheta)] (\epsilon_{t-j}^2 - \sigma^2), & \text{if } j = j'; \\ [q'_j(\vartheta)r_{j'}(\vartheta) + q_j(\vartheta)r'_{j'}(\vartheta)] \epsilon_{t-j}\epsilon_{t-j'}, & \text{otherwise.} \end{cases} \quad (\text{B.27})$$

Thus,

$$\begin{aligned} H_T(\theta') - H_T(\theta) &= T^{-1/2}(\theta' - \theta) \sum_{t=1}^T \sum_{j,j' \in \mathbb{N}} s'_{t,j,j'}(\vartheta_{\theta,\theta',t,j,j'}) \\ &= (\theta' - \theta) \sum_{j,j' \in \mathbb{N}} \left(T^{-1/2} \sum_{t=1}^T s'_{t,j,j'}(\vartheta_{\theta,\theta',t,j,j'}) \right), \end{aligned} \quad (\text{B.28})$$

which implies that

$$|H_T(\theta') - H_T(\theta)| \leq |\theta' - \theta| \sum_{j,j' \in \mathbb{N}} \left| T^{-1/2} \sum_{t=1}^T s'_{t,j,j'}(\vartheta_{\theta,\theta',t,j,j'}) \right|. \quad (\text{B.29})$$

Let $\sup_{\theta \in \Theta} |q_j(\theta)| = M_{q,j}$, $\sup_{\theta \in \Theta} |r_j(\theta)| = M_{r,j}$, $\sup_{\theta \in \Theta} |q'_j(\theta)| = M'_{q,j}$, $\sup_{\theta \in \Theta} |r'_j(\theta)| = M'_{r,j}$

for each $j \in \mathbb{N}$, then

$$\left| T^{-1/2} \sum_{t=1}^T s'_{t,j,j'}(\vartheta_{\theta,\theta',t,j,j'}) \right| \leq |S_{T,j,j'}|, \quad (\text{B.30})$$

where

$$S_{T,j,j'} = \begin{cases} (M'_{q,j}M_{r,j'} + M_{q,j}M'_{r,j'}) \left(T^{-1/2} \sum_{t=1}^T (\epsilon_{t-j}^2 - \sigma^2) \right), & \text{if } j = j'; \\ (M'_{q,j}M_{r,j'} + M_{q,j}M'_{r,j'}) \left(T^{-1/2} \sum_{t=1}^T \epsilon_{t-j}\epsilon_{t-j'} \right), & \text{otherwise.} \end{cases} \quad (\text{B.31})$$

As $E(S_{T,j,j'}) = 0$, let $\varsigma \triangleq \max\{\sigma_4 - \sigma^4, \sigma^4\}$ and we have

$$E(S_{T,j,j'}^2) = \text{Var}(S_{T,j,j'}) \leq (M'_{q,j}M_{r,j'} + M_{q,j}M'_{r,j'})^2 \varsigma \quad (\text{B.32})$$

for any $T \in \mathbb{Z}_+$ and $j, j' \in \mathbb{N}$. By *Hölder's inequality*, we have

$$E(|S_{T,j,j'}|) \leq E(S_{T,j,j'}^2)^{1/2} \leq (M'_{q,j}M_{r,j} + M_{q,j}M'_{r,j})\sqrt{\varsigma}.$$

Thus,

$$\begin{aligned}
& E \left(\sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| \right) \\
& \leq \sum_{j,j' \in \mathbb{N}} (M'_{q,j} M_{r,j'} + M_{q,j} M'_{r,j'}) \sqrt{\zeta} \\
& = [(\sum_{j \in \mathbb{N}} M'_{q,j})(\sum_{j' \in \mathbb{N}} M_{r,j'}) + (\sum_{j \in \mathbb{N}} M_{q,j})(\sum_{j' \in \mathbb{N}} M'_{r,j'})] \sqrt{\zeta} \\
& = (M'_q M_r + M_q M'_r) \sqrt{\zeta}.
\end{aligned} \tag{B.33}$$

Now for any $r > 0$, by (B.29) and (B.30), we have

$$\begin{aligned}
& Pr \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, r)} |H_T(\theta') - H_T(\theta)| > \epsilon \right) \\
& \leq Pr \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, r)} T^{a+1/2} |\theta' - \theta| \sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| > \epsilon \right) \\
& \leq Pr \left(\sup_{\theta \in \Theta} r \sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| > \epsilon \right) \\
& \leq Pr \left(\sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| > \frac{\epsilon}{r} \right).
\end{aligned}$$

By *Markov's inequality* and (B.33), we have

$$Pr \left(\sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| > \frac{\epsilon}{r} \right) \leq \frac{E \left(\sum_{j,j' \in \mathbb{N}} |S_{T,j,j'}| \right)}{\frac{\epsilon}{r}} \leq \frac{r(M'_q M_r + M_q M'_r) \sqrt{\zeta}}{\epsilon}.$$

Now take some $r < \frac{\epsilon^2}{(M'_q M_r + M_q M'_r) \sqrt{\zeta}}$, and we have

$$\limsup_{T \rightarrow \infty} Pr \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, r)} |H_T(\theta') - H_T(\theta)| > \epsilon \right) < \epsilon.$$

Thus, the sequence $\{H_T(\theta)\}_{T \in \mathbb{Z}_+}$ is stochastically equicontinuous, and the proof is complete. \square

Proof of Lemma 3.2.8. It is obvious that $A(-\theta)$, $A(-\theta_1)$, ..., $A(-\theta_m)$, and 1 satisfy the conditions for $Q(\theta)$.⁴³ It is clear that if $Q_1(\theta)$ and $Q_2(\theta)$ satisfy the conditions for $Q(\theta)$, then for any $a, b \in \mathbb{R}$, $aQ_1(\theta) + bQ_2(\theta)$ still satisfies the condition for $Q(\theta)$. Then, in order to show that the claim is true for any polynomial $\Psi_1(X_0, X_1, \dots, X_m)$, we only need

⁴³ $A(-\theta_1)$ and 1 are viewed as constant functions of θ .

to show that for any monomial of the form $X_0^{e_0} X_1^{e_1} \cdots X_m^{e_m}$ with $e_i \geq 0$ ($i = 0, 1, \dots, m$), $A^{e_0}(-\theta)A^{e_1}(-\theta_1) \cdots A^{e_m}(-\theta_m)$ satisfies the conditions for $Q(\theta)$.

For $j \in \mathbb{Z}_+$, select a specific $B_j(\theta)$ from the set $\{A(-\theta), A(-\theta_1), \dots, A(-\theta_m)\}$, where $A(-\theta_i)$ is viewed as a constant function of θ for $i = 1, \dots, m$. We know that for $Q(\theta) = \prod_{j=1}^k B_j(\theta)$ satisfies the conditions in Lemma 3.2.6 for $k = 0$ and 1. Let

$$\theta_T^* \triangleq \max\{|\theta_-|, |\theta^-|\} < 1.$$

We have the following claim.

Claim. Let $Q(\theta) = \prod_{j=1}^k B_j(\theta)$ for some $k \in \mathbb{Z}_+$, then $\sup_{\theta \in \Theta} |q_l(\theta)| \leq (l+1)^{k-1} \theta^{*l}$ and $\sup_{\theta \in \Theta} |q'_l(\theta)| \leq l(l+1)^{k-1} \theta^{*(l-1)}$ for any $l \in \mathbb{N}$.

We now prove the claim by induction. When $k = 1$, we have $\sup_{\theta \in \Theta} |q_l(\theta)| \leq \theta^{*l}$ and $\sup_{\theta \in \Theta} |q'_l(\theta)| \leq l \theta^{*(l-1)}$ for any $l \in \mathbb{N}$. Thus, the claim is true. Now suppose the claim is true for $k = p$. For $k = p+1$, we have $Q(\theta) = \left[\prod_{j=1}^p B_j(\theta) \right] B_{p+1}(\theta)$. Let $P(\theta) = \prod_{j=1}^p B_j(\theta) = \sum_{l=0}^{\infty} p_l(\theta) L^l$ and $B_{p+1}(\theta) = \sum_{l=0}^{\infty} b_l(\theta) L^l$, then

$$|q_l(\theta)| = \left| \sum_{u=0}^l p_u(\theta) b_{l-u}(\theta) \right| \leq \sum_{u=0}^l |p_u(\theta) b_{l-u}(\theta)| \leq \sum_{u=0}^l (u+1)^{p-1} \theta^{*u} \theta^{*(l-u)} \leq (l+1)^p \theta^{*l} \quad (\text{B.34})$$

and

$$|q'_l(\theta)| = l(l+1)^p \theta^{*(l-1)}, \quad (\text{B.35})$$

i.e., the claim is true for $k = p+1$. Thus, the claim is true for any $k \in \mathbb{Z}_+$.

By the above claim, it is clear that for any monomial of the form $X_0^{e_0} X_1^{e_1} \cdots X_m^{e_m}$, power series $A^{e_0}(-\theta)A^{e_1}(-\theta_1) \cdots A^{e_m}(-\theta_m)$ satisfies the conditions for $Q(\theta)$. The proof for $R(\theta)$ is the same, and the result is proved. \square

Proof of Lemma 3.2.9. By Equation (3.62), we have

$$\begin{aligned}
& \frac{\partial F_T}{\partial \delta} \\
&= -2 \sum [(1 - \delta A(-\theta) + \delta_T A(\phi_T) - \delta \delta_T A(-\theta) A(\phi_T)) \epsilon_{t,T}] (P \epsilon_{t,T}) \\
&= -2 \{ \sum \epsilon_{t,T} (P \epsilon_{t,T}) + \delta_T \sum (A(\phi_T) \epsilon_{t,T}) (P \epsilon_{t,T}) - \delta [\sum (A(-\theta) \epsilon_{t,T}) (P \epsilon_{t,T}) \\
&\quad + \delta_T \sum (A(-\theta) A(\phi_T) \epsilon_{t,T}) (P \epsilon_{t,T})] \},
\end{aligned} \tag{B.36}$$

where P is defined in (3.63). The FOC $\frac{\partial \mathcal{L}_T}{\partial \delta} = 0$ gives the following linear equation in δ

$$\begin{aligned}
& T^{-1} [\sum \epsilon_{t,T} (P \epsilon_{t,T}) + \delta_T \sum (A(\phi_T) \epsilon_{t,T}) (P \epsilon_{t,T})] \\
&= \delta T^{-1} [\sum (A(-\theta) \epsilon_{t,T}) (P \epsilon_{t,T}) + \delta_T \sum (A(-\theta) A(\phi_T) \epsilon_{t,T}) (P \epsilon_{t,T})].
\end{aligned} \tag{B.37}$$

Now we compute the above equation term by term using Lemma (3.2.8), and we have

$$\begin{aligned}
& T^{-1} \sum (A(-\theta) \epsilon_{t,T}) (A(-\theta) \epsilon_{t,T}) \\
&= T^{-1} \sum [(\sum_{j=0}^{\infty} (-\theta)^j L^{j+1}) \epsilon_{t,T}] [(\sum_{j=0}^{\infty} (-\theta)^j L^{j+1}) \epsilon_{t,T}] \\
&= \sum_{j=0}^{\infty} (-\theta)^{2j} + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1-\theta^2} + O_{p.u.}(T^{-1/2}),
\end{aligned} \tag{B.38}$$

$$\begin{aligned}
& T^{-1} \sum (A(\phi_T) \epsilon_{t,T}) (A(-\theta) \epsilon_{t,T}) \\
&= T^{-1} \sum [(\sum_{j=0}^{\infty} (\phi_T)^j L^{j+1}) \epsilon_{t,T}] [(\sum_{j=0}^{\infty} (-\theta)^j L^{j+1}) \epsilon_{t,T}] \\
&= \sum_{j=0}^{\infty} (-\phi_T \theta)^j + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1+\phi_T \theta} + O_{p.u.}(T^{-1/2}),
\end{aligned} \tag{B.39}$$

and

$$\begin{aligned}
& T^{-1/2} \sum \epsilon_{t,T} (P \epsilon_{t,T}) \\
&= T^{-1/2} [\sum \epsilon_{t,T} (A(-\theta) \epsilon_{t,T}) + \delta_T \sum \epsilon_{t,T} (A(-\theta) A(\phi_T) \epsilon_{t,T})] \\
&= T^{-1/2} \sum \sum_{j=0}^{\infty} (-\theta)^j \epsilon_{t,T} \epsilon_{t-j-1,T} + \epsilon'_{1,T}, \\
&= C_T(\theta) + \epsilon'_{1,T},
\end{aligned} \tag{B.40}$$

where $\epsilon'_{1,T} = T^{-1/2} \delta_T \sum \epsilon_{t,T} (A(-\theta) A(\phi_T) \epsilon_{t,T}) = O_{p.u.}(|\delta_T|)$. When $-\theta \neq \phi_T$, by Lemma 3.2.2, we have

$$A(-\theta) A(\phi_T) = \frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T}. \quad (\text{B.41})$$

Then by Lemma (3.2.8), we have

$$\begin{aligned} & T^{-1} \sum (A(\phi_T) \epsilon_{t,T}) (A(-\theta) A(\phi_T) \epsilon_{t,T}) \\ &= T^{-1} \sum (A(\phi_T) \epsilon_{t,T}) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \\ &= E \left[(A(\phi_T) \epsilon_{t,T}) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \right] + O_{p.u.}(T^{-1/2}) \\ &= \frac{1}{\theta + \phi_T} E [(A(\phi_T) \epsilon_{t,T}) (\theta L A(-\theta) \epsilon_{t,T}) + (A(\phi_T) \epsilon_{t,T}) (\phi_T L A(\phi_T) \epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\ &= \frac{1}{\theta + \phi_T} \left(\frac{\phi_T \theta}{1 + \phi_T \theta} + \frac{\phi_T^2}{1 - \phi_T^2} \right) + O_{p.u.}(T^{-1/2}) \\ &= \frac{\phi_T}{(1 + \phi_T \theta)(1 - \phi_T^2)} + O_{p.u.}(T^{-1/2}), \end{aligned} \quad (\text{B.42})$$

$$\begin{aligned} & T^{-1} \sum (A(-\theta) \epsilon_{t,T}) (A(-\theta) A(\phi_T) \epsilon_{t,T}) \\ &= T^{-1} \sum (A(-\theta) \epsilon_{t,T}) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \\ &= E \left[(A(-\theta) \epsilon_{t,T}) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \right] + O_{p.u.}(T^{-1/2}) \\ &= \frac{1}{\theta + \phi_T} E [(A(-\theta) \epsilon_{t,T}) (\theta L A(-\theta) \epsilon_{t,T}) + (A(-\theta) \epsilon_{t,T}) (\phi_T L A(\phi_T) \epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\ &= \frac{1}{\theta + \phi_T} \left(\frac{-\theta^2}{1 - \theta^2} + \frac{-\phi_T \theta}{1 + \phi_T \theta} \right) + O_{p.u.}(T^{-1/2}) \\ &= \frac{-\theta}{(1 + \phi_T \theta)(1 - \theta^2)} + O_{p.u.}(T^{-1/2}), \end{aligned} \quad (\text{B.43})$$

and

$$\begin{aligned}
& T^{-1} \sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T}) \\
&= T^{-1} \sum \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \\
&= E \left[\left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \left(\frac{L[\theta A(-\theta) + \phi_T A(\phi_T)]}{\theta + \phi_T} \epsilon_{t,T} \right) \right] + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{(\theta + \phi_T)^2} E [(\theta LA(-\theta)\epsilon_{t,T})(\theta LA(-\theta)\epsilon_{t,T}) + (\theta LA(-\theta)\epsilon_{t,T})(\phi_T LA(\phi_T)\epsilon_{t,T}) \\
&\quad + (\phi_T LA(\phi_T)\epsilon_{t,T})(\theta LA(-\theta)\epsilon_{t,T}) + (\phi_T LA(\phi_T)\epsilon_{t,T})(\phi_T LA(\phi_T)\epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{(\theta + \phi_T)^2} \left(\frac{\theta^2}{1 - \theta^2} + \frac{2\phi_T\theta}{1 + \phi_T\theta} + \frac{\phi_T^2}{1 - \phi_T^2} \right) + O_{p.u.}(T^{-1/2}) \\
&= \frac{1 - \phi_T\theta}{(1 - \theta^2)(1 + \phi_T\theta)(1 - \phi_T^2)} + O_{p.u.}(T^{-1/2}).
\end{aligned} \tag{B.44}$$

When $-\theta = \phi_T$, we have

$$\begin{aligned}
& T^{-1} \sum (A(\phi_T)\epsilon_{t,T})(A(\phi_T)^2\epsilon_{t,T}) \\
&= E [(A(\phi_T)\epsilon_{t,T})(A(\phi_T)^2\epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\
&= \sum_{j=1}^{\infty} \phi_T^j (j\phi_T^{j-1}) + O_{p.u.}(T^{-1/2}) \\
&= \frac{\phi_T}{(1 - \phi_T^2)^2} + O_{p.u.}(T^{-1/2})
\end{aligned} \tag{B.45}$$

and

$$\begin{aligned}
& T^{-1} \sum (A(\phi_T)^2\epsilon_{t,T})(A(\phi_T)^2\epsilon_{t,T}) \\
&= E [(A(\phi_T)^2\epsilon_{t,T})(A(\phi_T)^2\epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\
&= \sum_{j=1}^{\infty} (j\phi_T^{j-1})^2 + O_{p.u.}(T^{-1/2}) \\
&= \frac{1 + \phi_T^2}{(1 - \phi_T^2)^3} + O_{p.u.}(T^{-1/2}).
\end{aligned} \tag{B.46}$$

We see that Formulae (B.42)–(B.44) match (B.45) and (B.46) when $-\theta = \phi_T$. Thus, we adopt

Formulae (B.42)–(B.44) no matter whether $-\theta = \phi_T$. Based on pervious calculation, we

have

$$\begin{aligned}
& T^{-1} [\sum (A(-\theta)\epsilon_{t,T})(P\epsilon_{t,T}) + \delta_T \sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(P\epsilon_{t,T})] \\
&= T^{-1} [\sum (A(-\theta)\epsilon_{t,T})(A(-\theta)(1 + \delta_T A(\phi_T))\epsilon_{t,T})] \\
&\quad + \delta_T T^{-1} [\sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(A(-\theta)(1 + \delta_T A(\phi_T))\epsilon_{t,T})] \\
&= T^{-1} [\sum (A(-\theta)\epsilon_{t,T})(A(-\theta)\epsilon_{t,T}) + \delta_T \sum (A(-\theta)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] \\
&\quad + \delta_T T^{-1} [\sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(A(-\theta)\epsilon_{t,T}) + \delta_T \sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] \\
&= E[(A(-\theta)\epsilon_{t,T})(A(-\theta)\epsilon_{t,T})] + 2\delta_T E[(A(-\theta)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] \\
&\quad + \delta_T^2 E[(A(-\theta)A(\phi_T)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1-\theta^2} + 2\delta_T \frac{-\theta}{(1+\phi_T\theta)(1-\theta^2)} + \delta_T^2 \frac{1-\phi_T\theta}{(1-\theta^2)(1+\phi_T\theta)(1-\phi_T^2)} + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1-\theta^2} \left[1 - \delta_T \frac{2\theta}{1+\phi_T\theta} + \delta_T^2 \frac{1-\phi_T\theta}{(1+\phi_T\theta)(1-\phi_T^2)} \right] + O_{p.u.}(T^{-1/2}),
\end{aligned} \tag{B.47}$$

and

$$\begin{aligned}
& T^{-1} \sum (A(\phi_T)\epsilon_{t,T})(P\epsilon_{t,T}) \\
&= T^{-1} [\sum (A(\phi_T)\epsilon_{t,T})(A(-\theta)\epsilon_{t,T}) + \delta_T \sum (A(\phi_T)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] \\
&= E[(A(\phi_T)\epsilon_{t,T})(A(-\theta)\epsilon_{t,T})] + \delta_T E[(A(\phi_T)\epsilon_{t,T})(A(-\theta)A(\phi_T)\epsilon_{t,T})] + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1+\phi_T\theta} + \delta_T \frac{\phi_T}{(1+\phi_T\theta)(1-\phi_T^2)} + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1+\phi_T\theta} \left(1 + \delta_T \frac{\phi_T}{1-\phi_T^2} \right) + O_{p.u.}(T^{-1/2}) \\
&= \frac{1}{1+\phi_T\theta} \frac{1+\phi_T\theta}{1-\phi_T^2} + O_{p.u.}(T^{-1/2}).
\end{aligned} \tag{B.48}$$

Thus, ignoring a set of measure zero (or measure zero asymptotically) on which the de-

nominator takes value 0,⁴⁴ Equation (B.37) becomes

$$\begin{aligned}
\delta &= \frac{T^{-1/2}C_T(\theta) + O_{p.u.}(T^{-1/2}|\delta_T|) + \delta_T \left[\frac{1}{1+\phi_T\theta} \frac{1+\phi_T\theta_T}{1-\phi_T^2} + O_{p.u.}(T^{-1/2}) \right]}{\frac{1}{1-\theta^2} \left[1 - \delta_T \frac{2\theta}{1+\phi_T\theta} + \delta_T^2 \frac{1-\phi_T\theta}{(1+\phi_T\theta)(1-\phi_T^2)} \right] + O_{p.u.}(T^{-1/2})} \\
&= \frac{T^{-1/2}(1-\theta^2)C_T(\theta) + \delta_T \left(\frac{1-\theta^2}{1+\phi_T\theta} \frac{1+\phi_T\theta_T}{1-\phi_T^2} \right) + O_{p.u.}(T^{-1/2}|\delta_T|)}{1 - \delta_T \frac{2\theta}{1+\phi_T\theta} + \delta_T^2 \frac{1-\phi_T\theta}{(1+\phi_T\theta)(1-\phi_T^2)} + O_{p.u.}(T^{-1/2})} \\
&= \left[T^{-1/2}(1-\theta^2)C_T(\theta) + \delta_T \left(\frac{1-\theta^2}{1+\phi_T\theta} \frac{1+\phi_T\theta_T}{1-\phi_T^2} \right) \right] \left[1 - \delta_T \frac{2\theta}{1+\phi_T\theta} + \delta_T^2 \frac{1-\phi_T\theta}{(1+\phi_T\theta)(1-\phi_T^2)} \right]^{-1} \\
&\quad + O_{p.u.}(T^{-1}) + O_{p.u.}(T^{-1/2}|\delta_T|).
\end{aligned} \tag{B.50}$$

It is clear that

$$f_1(\theta, \delta_T, \phi_T) = 1 + O(|\delta_T|). \tag{B.51}$$

Thus, we have⁴⁵

$$\delta = T^{-1/2}(1-\theta^2)C_T(\theta) + \delta_T \left(\frac{1-\theta^2}{1+\phi_T\theta} \frac{1+\phi_T\theta_T}{1-\phi_T^2} \right) f_1(\theta, \delta_T, \phi_T) + O_{p.u.}(\max\{T^{-1}, T^{-1/2}|\delta_T|\}).$$

Thus, the proof is complete. \square

Proof of Theorem 3.2.11. As the estimate $\hat{\xi}_T$ satisfies the FOC $\frac{\partial F_T}{\partial \delta} = 0$, Lemma 3.2.9 implies that

$$\Delta \hat{\delta}_T = \left(\frac{1 - \hat{\theta}_T^2}{1 + \phi_T \hat{\theta}_T} \frac{1 + \phi_T \theta_T}{1 - \phi_T^2} f_1(\hat{\theta}_T, \delta_T, \phi_T) - 1 \right) \delta_T + T^{-1/2}(1 - \hat{\theta}_T^2)C_T(\hat{\theta}_T) + \varepsilon_{2,T}, \tag{B.52}$$

⁴⁴If the innovation $\epsilon_{t,T}$ is continuously distributed over \mathbb{R} , then it is clear that for a given θ , the set on which the denominator takes value 0 is of measure zero. If $\epsilon_{t,T}$ has a *point mass*, then the set on which the denominator takes value 0 may be of some positive measure which will converge to zero as $T \rightarrow \infty$. Note that if the denominator takes value 0 for the given θ , then we have the “rare” case that $F_T(\xi)$ is flat in the direction of δ for that θ .

Thus, a more direct treatment for this issue is to define

$$S_T^\delta(\theta_-, \theta^-, \xi_T) \triangleq \left\{ \omega \in \Omega \left| \inf_{\theta \in \Theta} \left(\sum (A(-\theta)\epsilon_{t,T})(P\epsilon_{t,T}) + \delta_T \sum (A(-\theta)A(\phi_T)\epsilon_{t,T})(P\epsilon_{t,T}) \right) > 0 \right. \right\}, \tag{B.49}$$

and consider $\omega \in S_T^\delta(\theta_-, \theta^-, \xi_T)$ only. (By Lemma 3.2.8, $\lim_{T \rightarrow \infty} Pr(S_T^\delta(\theta_-, \theta^-, \xi_T)) = 1$.)

⁴⁵In this paper, we use notations, like $O(T^a)$, $O_p(T^a)$, $o_{p.u.}(T^a)$, etc., to replace the corresponding *higher-order infinitesimals* for notational simplicity. The use of these notations does not mean general identities like $O_{p.u.}(T^{-1}) + O_{p.u.}(T^{-1/2}|\delta_T|) = O_{p.u.}(\max\{T^{-1}, T^{-1/2}|\delta_T|\})$.

and

$$\begin{aligned}
& \frac{1-\hat{\theta}_T^2}{1+\phi_T\hat{\theta}_T} \frac{1+\phi_T\theta_T}{1-\phi_T^2} f_1(\hat{\theta}_T, \delta_T, \phi_T) - 1 \\
&= f_1(\hat{\theta}_T, \delta_T, \phi_T) \left[\frac{1-\hat{\theta}_T^2}{1+\phi_T\hat{\theta}_T} \frac{1+\phi_T\theta_T}{1-\phi_T^2} - \left(1 - \delta_T \frac{2\hat{\theta}_T}{1+\phi_T\hat{\theta}_T} + \delta_T^2 \frac{1-\phi_T\hat{\theta}_T}{(1+\phi_T\hat{\theta}_T)(1-\phi_T^2)} \right) \right] \\
&= \frac{f_1(\hat{\theta}_T, \delta_T, \phi_T)}{(1+\phi_T\hat{\theta}_T)(1-\phi_T^2)} [(1-\hat{\theta}_T^2)(1+\phi_T\theta_T) - (1+\phi_T\hat{\theta}_T)(1-\phi_T^2) + 2\delta_T\hat{\theta}_T(1-\phi_T^2) \\
&\quad - \delta_T^2(1-\phi_T\hat{\theta}_T)] \\
&= \frac{f_1(\hat{\theta}_T, \delta_T, \phi_T)}{(1+\phi_T\hat{\theta}_T)(1-\phi_T^2)} f_2(\hat{\theta}_T, \delta_T, \phi_T) \triangle \hat{\theta}_T.
\end{aligned} \tag{B.53}$$

Thus,

$$\triangle \hat{\delta}_T = \frac{f_1(\hat{\theta}_T, \delta_T, \phi_T)}{(1+\phi_T\hat{\theta}_T)(1-\phi_T^2)} f_2(\hat{\theta}_T, \delta_T, \phi_T) \triangle \hat{\theta}_T \delta_T + T^{-1/2}(1-\hat{\theta}_T^2)C_T(\hat{\theta}_T) + \varepsilon_{2,T}. \tag{B.54}$$

The corollary can be implied from the above formula readily. Thus, the proof is complete. \square

Proof of Lemma 3.2.12. Similar to the proof of Lemma 3.2.9, we calculate the terms in (3.83) first. By Lemma 3.2.3 and 3.2.8, we have

$$\begin{aligned}
& T^{-1/2} \sum (A(-\theta)\epsilon_{t,T})(A(-\theta)^2\epsilon_{t,T}) \\
&= T^{-1/2} \sum [(\sum_{j=1}^{\infty} (-\theta)^{j-1} L^j) \epsilon_{t,T}] [(\sum_{j=2}^{\infty} (j-1)(-\theta)^{j-2} L^j) \epsilon_{t,T}] \\
&= T^{1/2} \sum \sum_{j=2}^{\infty} [(j-1)(-\theta)^{2j-3}] + O_{p.u.}(1) \\
&= T^{1/2} \frac{-\theta}{(1-\theta^2)^2} + O_{p.u.}(1),
\end{aligned} \tag{B.55}$$

$$\begin{aligned}
& T^{-1/2} \sum (A(\phi_T)\epsilon_{t,T})(A(-\theta)^2\epsilon_{t,T}) \\
&= T^{-1/2} \sum [(\sum_{j=1}^{\infty} \phi_T^{j-1} L^j) \epsilon_{t,T}] [(\sum_{j=2}^{\infty} (j-1)(-\theta)^{j-2} L^j) \epsilon_{t,T}] \\
&= T^{1/2} \sum \sum_{j=2}^{\infty} [\phi_T^{j-1} (j-1)(-\theta)^{j-2}] + O_{p.u.}(1) \\
&= T^{1/2} \frac{\phi_T}{(1+\phi_T\theta)^2} + O_{p.u.}(1),
\end{aligned} \tag{B.56}$$

$$T^{-1/2} \sum (A(-\theta)\epsilon_{t,T})(A(-\theta)^2 A(\phi_T)\epsilon_{t,T}) = T^{1/2} g_1(\theta, \phi_T) + O_{p.u.}(1), \tag{B.57}$$

$$T^{-1/2} \sum (A(\phi_T) \epsilon_{t,T}) (A(-\theta)^2 A(\phi_T) \epsilon_{t,T}) = T^{1/2} g_2(\theta, \phi_T) + O_{p.u.}(1), \quad (\text{B.58})$$

$$T^{-1/2} \sum (A(-\theta) A(\phi_T) \epsilon_{t,T}) (A(-\theta)^2 \epsilon_{t,T}) = T^{1/2} g_3(\theta, \phi_T) + O_{p.u.}(1), \quad (\text{B.59})$$

and

$$T^{-1/2} \sum (A(-\theta) A(\phi_T) \epsilon_{t,T}) (A(-\theta)^2 A(\phi_T) \epsilon_{t,T}) = T^{1/2} g_4(\theta, \phi_T) + O_{p.u.}(1), \quad (\text{B.60})$$

where $g_1(\theta, \phi_T) = \frac{\theta^2}{(1+\phi_T\theta)(1-\theta^2)^2}$, $g_2(\theta, \phi_T) = \frac{\phi_T^2}{(1+\phi_T\theta)^2(1-\phi_T^2)}$, $g_3(\theta, \phi_T) = \frac{1+\phi_T\theta^3}{(1-\theta^2)^2(1+\phi_T\theta)^2}$, and $g_4(\theta, \phi_T) = \frac{\phi_T - \theta - \phi_T\theta^2 + \phi_T^2\theta^3}{(1-\phi_T^2)(1-\theta^2)^2(1+\phi_T\theta)^2}$. It is clear that $\frac{\partial g_1}{\partial \theta}$, $\frac{\partial g_2}{\partial \theta}$, $\frac{\partial g_3}{\partial \theta}$, and $\frac{\partial g_4}{\partial \theta}$ are all continuous and bounded functions.

Thus,

$$\begin{aligned} & T^{-1/2} \sum [(-\delta A(-\theta) + \delta_T A(\phi_T)) \epsilon_{t,T}] (A(-\theta)^2 \epsilon_{t,T}) \\ &= T^{1/2} \left[\frac{\delta \theta}{(1-\theta^2)^2} + \frac{\delta_T \phi_T}{(1+\phi_T\theta)^2} \right] + O_{p.u.}(\max\{|\delta|, |\delta_T|\}). \end{aligned} \quad (\text{B.61})$$

By Lemma 3.2.8, we have

$$\begin{aligned} & \frac{\partial F_T}{\partial \theta} \\ &= 2\delta \left\{ \sum \epsilon_{t,T} (A(-\theta)^2 \epsilon_{t,T}) + \delta_T \sum \epsilon_{t,T} (A(-\theta)^2 A(\phi_T) \epsilon_{t,T}) \right. \\ & \quad \left. + \sum [(-\delta A(-\theta) + \delta_T A(\phi_T) - \delta \delta_T A(-\theta) A(\phi_T)) \epsilon_{t,T}] [A(-\theta)^2 (1 + \delta_T A(\phi_T)) \epsilon_{t,T}] \right\} \\ &= 2\delta \left\{ T^{1/2} D_T(\theta) + O_{p.u.}(T^{1/2} |\delta_T|) + T \left[\frac{\delta \theta}{(1-\theta^2)^2} + \frac{\delta_T \phi_T}{(1+\phi_T\theta)^2} \right] + O_{p.u.}(T^{1/2} \max\{|\delta|, |\delta_T|\}) \right. \\ & \quad \left. + \sum [(-\delta A(-\theta) + \delta_T A(\phi_T)) \epsilon_{t,T}] (A(-\theta)^2 \delta_T A(\phi_T) \epsilon_{t,T}) \right. \\ & \quad \left. + \sum (-\delta \delta_T A(-\theta) A(\phi_T) \epsilon_{t,T}) [A(-\theta)^2 (1 + \delta_T A(\phi_T)) \epsilon_{t,T}] \right\} \\ &= 2\delta \left\{ T^{1/2} D_T(\theta) + T \left[\frac{\delta \theta}{(1-\theta^2)^2} + \frac{\delta_T \phi_T}{(1+\phi_T\theta)^2} \right] + O_{p.u.}(T^{1/2} \max\{|\delta|, |\delta_T|\}) - \delta \delta_T T g_1(\theta, \phi_T) \right. \\ & \quad \left. + \delta_T^2 T g_2(\theta, \phi_T) - \delta \delta_T T g_3(\theta, \phi_T) - \delta \delta_T^2 T g_4(\theta, \phi_T) \right\} \\ &= 2\delta T^{1/2} \left\{ D_T(\theta) + T^{1/2} \left[\frac{\delta \theta}{(1-\theta^2)^2} + \delta_T \left(\frac{\phi_T}{(1+\phi_T\theta)^2} + f_3(\xi, \delta_T, \phi_T) \right) \right] + O_{p.u.}(\max\{|\delta|, |\delta_T|\}) \right\}, \end{aligned} \quad (\text{B.62})$$

where

$$\begin{aligned} f_3(\xi, \delta_T, \phi_T) &= -\delta g_1(\theta, \phi_T) + \delta_T g_2(\theta, \phi_T) - \delta g_3(\theta, \phi_T) - \delta \delta_T g_4(\theta, \phi_T) \\ &= \frac{-\delta \theta^2}{(1+\phi_T \theta)(1-\theta^2)^2} + \frac{\delta_T \phi_T^2}{(1+\phi_T \theta)^2(1-\phi_T^2)} - \frac{\delta(1+\phi_T \theta^3)}{(1-\theta^2)^2(1+\phi_T \theta)^2} - \frac{\delta \delta_T(\phi_T - \theta - \phi_T \theta^2 + \phi_T^2 \theta^3)}{(1-\phi_T^2)(1-\theta^2)^2(1+\phi_T \theta)^2}. \end{aligned}$$

Thus, the proof is complete. \square

Proof of Lemma 3.2.14. The first equality follows from (3.72), and the second equality follows from Lemma 3.2.9 and 3.2.12. \square

Proof of Theorem 3.1.2. (1) By Lemmas 3.2.14, 3.2.9, and 3.2.12, we know that for $1/2\text{-}\langle 0 \rangle$ sequences

$$\mathcal{V}'_T(\theta) - \mathcal{S}_T(\theta|d_T, \phi_T) = O_{p.u.}(T^{-1/2}), \quad (\text{B.63})$$

and

$$\mathcal{S}_T(\theta|d_T, \phi_T) - \mathcal{S}_T(\theta|0, \phi_T) = O_{p.u.}(|d_T|), \quad (\text{B.64})$$

where $\mathcal{S}_T(\theta|0, \phi_T)$ is given in (3.3). Thus, the difference between $\mathcal{V}'_T(\theta)$ and the shape analysis function $\mathcal{S}_T(\theta|0, \phi_T)$ satisfies

$$\mathcal{V}'_T(\theta) - \mathcal{S}_T(\theta|0, \phi_T) = O_{p.u.}(\max\{T^{-1/2}, |d_T|\}), \quad (\text{B.65})$$

Note that the shape analysis function (3.3) doesn't depend on d_T and ϕ_T . As the MLE $\hat{\theta}_T$ is determined by the behavior of $\mathcal{V}'_T(\theta)$ on (θ_-, θ^-) , i.e., by the sample path $\mathcal{V}'_{\omega, T}$ associated with ω on (θ_-, θ^-) , (3.3) implies that θ_T is asymptotically unidentified in this case.

By (3.96)–(3.98), we know that $V(\theta)$ are nonzero matrices for all the θ 's, and it is clear that the MLE $\hat{\theta}_T$ doesn't converge to θ_T in probability. Namely, $\hat{\theta}_T$ is not consistent. It is clear that for all the $1/2\text{-}\langle 0 \rangle$ sequences, the asymptotic distribution is the same and is determined by the shape analysis function (3.3) and (3.73).

(2) For a $1/2\text{-}\langle 1 \rangle$ sequences with $\lim_{T \rightarrow \infty} d_T = d_0 \in \mathbb{R} \setminus \{0\}$, we know that (B.63) also holds, and

$$\mathcal{S}_T(\theta|d_T, \phi_T) - \mathcal{S}_T(\theta|d_0, \phi_0) = O_{p.u.}(|d_T - d_0|), \quad (\text{B.66})$$

where $\mathcal{S}_T(\theta|d_0, \phi_0)$ is given in (3.4), which determines the asymptotic distribution. It is clear that the asymptotic distribution of $\hat{\xi}_T$ depends on the value of ϕ_T , as d_0 is nonzero. Thus, $\hat{\theta}_T$ is asymptotically identified. The only terms that depend on T in the expression of $\mathcal{S}_T(\theta|d_0, \phi_0)$ are GP's C_T and D_T , and it is clear that $\hat{\theta}_T$ will not converge to θ_T in probability by (3.96)–(3.98). Namely, $\hat{\theta}_T$ is not consistent.

(3) For a $1/2\text{-}\langle \infty \rangle$ sequences, let's consider the case that $\lim_{T \rightarrow \infty} d_T = +\infty$, and the case that $\lim_{T \rightarrow \infty} d_T = -\infty$ can be dealt similarly. WLOG, we assume that d_T are positive for all $T \geq 1$. By Lemma 3.2.8, (3.92) implies that

$$\frac{\mathcal{W}_{1,T}(\theta)}{d_T} - \frac{1 - \theta^2}{1 + \phi_T \theta} = O_{p.u.}(\max\{d_T^{-1}, |\delta_T|\}), \quad (\text{B.67})$$

and note that

$$\frac{1 - \theta^2}{1 + \phi_T \theta} \geq c \text{ over } \Theta \text{ for some } c > 0. \quad (\text{B.68})$$

By Lemma 3.2.14, we have

$$\mathcal{W}_{2,T}(\theta) = G_T(\theta) + \frac{T^{1/2}\theta\varepsilon_{1,T}}{1-\theta^2} + O_{p.u.}(\max\{|f_T(\theta)|, |\delta_T|\}) + d_T f_5(c_T(\theta), \delta_T, \phi_T), \quad (\text{B.69})$$

where f_5 is defined in (3.95). By Theorem 3.2.11 and definition of f_3 , it is clear that

$$f_3(c_T(\theta), \delta_T, \phi_T) = f_3((\delta_T, \theta)', \delta_T, \phi_T) + f_6(\theta, \delta_T, \phi_T), \quad (\text{B.70})$$

by Theorem 3.2.11, where

$$\begin{aligned}
f_6(\theta, \delta_T, \phi_T) &= \left(\frac{-\theta^2}{(1+\phi_T\theta)(1-\theta^2)^2} - \frac{1+\phi_T\theta^3}{(1-\theta^2)^2(1+\phi_T\theta)^2} - \frac{\delta_T(\phi_T-\theta-\phi_T\theta^2+\phi_T^2\theta^3)}{(1-\phi_T^2)(1-\theta^2)^2(1+\phi_T\theta)^2} \right) \\
&\quad \cdot \left(\frac{f_1(\theta, \delta_T, \phi_T)}{(1+\phi_T\theta)(1-\phi_T^2)} f_2(\theta, \delta_T, \phi_T)(\theta - \theta_T)\delta_T + T^{-1/2}(1-\theta^2)C_T(\theta) + \varepsilon_{1,T} \right) \\
&= f_7(\theta, \delta_T, \phi_T) + \varepsilon_{3,T},
\end{aligned} \tag{B.71}$$

where

$$\begin{aligned}
&f_7(\theta, \delta_T, \phi_T) \\
&= \left(\frac{-\theta^2}{(1+\phi_T\theta)(1-\theta^2)^2} - \frac{1+\phi_T\theta^3}{(1-\theta^2)^2(1+\phi_T\theta)^2} - \frac{\delta_T(\phi_T-\theta-\phi_T\theta^2+\phi_T^2\theta^3)}{(1-\phi_T^2)(1-\theta^2)^2(1+\phi_T\theta)^2} \right) \frac{f_1(\theta, \delta_T, \phi_T)}{(1+\phi_T\theta)(1-\phi_T^2)} f_2(\theta, \delta_T, \phi_T)(\theta - \theta_T)\delta_T
\end{aligned} \tag{B.72}$$

and

$$\varepsilon_{3,T} = O_{p.u.}(T^{-1/2}).$$

Thus, by (3.95), (B.70), (B.71), and (B.72), (B.69) becomes

$$\mathcal{W}_{2,T}(\theta) = G_T(\theta) + d_T f_8(\theta, \delta_T, \phi_T) + O_{p.u.}(|\delta_T|), \tag{B.73}$$

where

$$f_8(\theta, \delta_T, \phi_T) = f_5((\delta_T, \theta)', \delta_T, \phi_T) + f_7(\theta, \delta_T, \phi_T). \tag{B.74}$$

Function f_8 looks complicated. However, we know that

$$\begin{aligned}
&f_8(\theta, \delta_T, \phi_T) - h(\theta|\phi_T) \\
&= \frac{1}{1+\phi_T\theta} \frac{\theta}{1-\theta^2} \left[\left(1 + \frac{\delta_T\phi_T}{1-\phi_T^2} \right) f_1(\theta, \delta_T, \phi_T) - 1 \right] + f_3((\delta_T, \theta)', \delta_T, \phi_T) + f_7(\theta, \delta_T, \phi_T) \\
&= O(|\delta_T|).
\end{aligned} \tag{B.75}$$

Thus, function $f_8(\theta, \delta_T, \phi_T)$ converges to $h(\theta|\phi_T)$ uniformly over Θ , as $T \rightarrow \infty$. Moreover, it is clear that

$$\frac{\partial}{\partial \theta}(f_8(\theta, \delta_T, \phi_T) - h(\theta|\phi_T)) = O(|\delta_T|). \tag{B.76}$$

By (3.108), Lemma 3.2.8, (B.73), and (B.75), we have

$$\frac{\mathcal{W}_{2,T}(\theta)}{d_T} - h(\theta|\phi_T) = O_{p.u.}(\max\{d_T^{-1}, |\delta_T|\}). \quad (\text{B.77})$$

Given $\phi_T \in (\theta_-, \theta^-)$, it is clear that

$$h(\theta_-|\phi_T) < 0 \text{ and } h(\theta^-|\phi_T) > 0. \quad (\text{B.78})$$

The graphs of function $h(\theta|\phi)$ for various values of ϕ are depicted in Figure B.1.

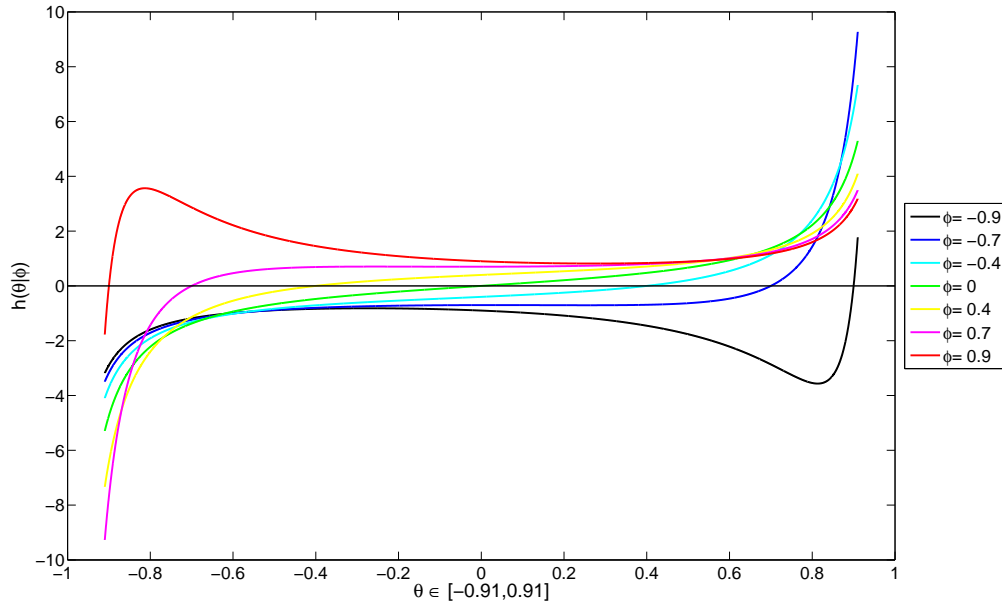


Figure B.1: Graphs of $h(\theta|\phi)$ for Various ϕ

Let

$$S_{\mathcal{W}_{1,T}}(\theta_-, \theta^-, \xi_T) \triangleq \{\omega \in \Omega \mid \inf_{\theta \in \Theta} \mathcal{W}_{1,T}(\theta) > 0\} \cap S_T^\delta(\theta_-, \theta^-, \xi_T),$$

$$S_{\mathcal{W}_{2,T}}(\theta_-, \theta^-, \xi_T) \triangleq \{\omega \in \Omega \mid \mathcal{V}'_T(\theta_-) < 0 \text{ and } \mathcal{V}'_T(\theta^-) > 0\} \cap S_T^\delta(\theta_-, \theta^-, \xi_T),$$

and

$$S_T^{\text{int}}(\theta_-, \theta^-, \xi_T) \triangleq S_{\mathcal{W}_{1,T}}(\theta_-, \theta^-, \xi_T) \cap S_{\mathcal{W}_{2,T}}(\theta_-, \theta^-, \xi_T),$$

where $S_T^{\delta}(\theta_-, \theta^-, \xi_T)$ is defined in (B.49), then for every $\omega \in S_T^{\text{int}}(\theta_-, \theta^-, \xi_T)$, the minima of $\mathcal{V}_T(\theta)$ occur in the interior of Θ , which implies that the FOC $\mathcal{V}'_T(\theta) = 0$ becomes a necessary condition for minima. By (B.67) and (B.68),

$$Pr(S_{\mathcal{W}_{1,T}}(\theta_-, \theta^-, \xi_T)) \rightarrow 1, \text{ as } T \rightarrow \infty.$$

By (B.77) and (B.78),

$$Pr(S_{\mathcal{W}_{2,T}}(\theta_-, \theta^-, \xi_T)) \rightarrow 1, \text{ as } T \rightarrow \infty.$$

Thus,

$$Pr(S_T^{\text{int}}(\theta_-, \theta^-, \xi_T)) \rightarrow 1, \text{ as } T \rightarrow \infty. \quad (\text{B.79})$$

Namely, the probability that the minima of $\mathcal{V}_T(\theta)$ occur in the interior of Θ approaches 1, as $T \rightarrow \infty$.

On set $S_T^{\text{int}}(\theta_-, \theta^-, \xi_T)$, by Lemma (3.2.14), the FOC $\mathcal{V}'_T(\theta) = 0$ becomes $\mathcal{W}_{2,T}(\theta) = 0$, and (B.73) becomes

$$-G_T(\theta) - \varepsilon_{4,T} = d_T f_8(\theta, \delta_T, \phi_T), \quad (\text{B.80})$$

where

$$\varepsilon_{4,T} = O_{p.u}(|\delta_T|). \quad (\text{B.81})$$

Note that the equation $h(\theta|\phi) = 0$ has the unique solution $\theta = -\phi$, and

$$h'(\theta|\phi) = \frac{(1 + \phi\theta)(1 - \theta^2) - \theta[\phi(1 - \theta^2) + (1 + \phi\theta)(-2\theta)]}{(1 + \phi\theta)^2(1 - \theta^2)^2} - \frac{2\phi^2}{(1 + \phi\theta)^3}, \quad (\text{B.82})$$

which gives

$$h'(-\phi|\phi) = \frac{1}{(1-\phi^2)^3} \geq 1. \quad (\text{B.83})$$

Furthermore, it is clear that $h(\theta|\phi_T)$ (as functions of θ) have the *strong upward transversal property*. That is, there exists $\varepsilon, r, T' > 0$, such that for every $T \geq T'$, (1) $h(\theta|\phi_T) < -\varepsilon$ for every $\theta < -\phi_0 - r$, (2) $h(\theta|\phi_T) > \varepsilon$ for every $\theta > -\phi_0 + r$, and (3) $h'(\theta|\phi_T) \geq c'$ over $[-\phi_0 - r, -\phi_0 + r]$ for some $c' > 0$.

Note that by Remark 3.2.15 and definition of f_7 , it is obvious that $f_8(\theta_T, \delta_T, \phi_T) = 0$. By (B.75) and (B.76), we know that $f_8(\theta, \delta_T, \phi_T)$ (as functions of θ) also have the strong upward transversal property. Namely, there exists $\varepsilon', r', T'' > 0$, such that for every $T \geq T''$, (1) $f_8(\theta, \delta_T, \phi_T) < -\varepsilon'$ for every $\theta < -\theta_0 - r'$, (2) $f_8(\theta, \delta_T, \phi_T) > \varepsilon'$ for every $\theta > -\theta_0 + r'$, and (3) $\frac{\partial}{\partial \theta} f_8(\theta, \delta_T, \phi_T) \geq c''$ over $[-\theta_0 - r', -\theta_0 + r']$ for some $c'' > 0$.

Let $\tilde{G}_T(\theta|\xi_T) \triangleq -G_T(\theta) - \varepsilon_{4,T}$, and define

$$\begin{aligned} & S_T^\circ(\theta_-, \theta^-, \xi_T) \\ & \triangleq \left\{ \omega \in S_T^{\text{int}}(\theta_-, \theta^-, \xi_T) \mid d_T^{-1} \left(\sup_{\theta \in \Theta} |\tilde{G}_T(\theta|\xi_T)| \right) \leq \varepsilon' \text{ and } \sup_{\theta \in \Theta} |\varepsilon_{4,T}| \leq \delta_T^{1/2} \right\}, \end{aligned}$$

then by strong upward transversal property, we know that for every $\omega \in S_T^\circ(\theta_-, \theta^-, \xi_T)$, Equation (B.80) has at least one solution θ_T^* , which must fall in the interval $[-\phi_0 - r', -\phi_0 + r']$. Moreover, it is clear that on $S_T^\circ(\theta_-, \theta^-, \xi_T)$,

$$\theta_T^* - \theta_T = O_p(d_T^{-1}). \quad (\text{B.84})$$

By Lemma 3.2.8, (B.81) and $\lim_{T \rightarrow \infty} d_T = +\infty$, we have

$$\lim_{T \rightarrow \infty} Pr(S_T^{\text{int}}(\theta_-, \theta^-, \xi_T) \setminus S_T^\circ(\theta_-, \theta^-, \xi_T)) = 0.$$

Then by (B.79), we have

$$\lim_{T \rightarrow \infty} \Pr(S_T^\circ(\theta_-, \theta^-, \xi_T)) = 1. \quad (\text{B.85})$$

As $\hat{\theta}_T$ must equal some θ_T^* , we have shown that $\hat{\theta}_T$ is consistent.

We further explore the asymptotic distribution π_{ξ_T} . It is clear that for (almost) every $\omega \in S_T^\circ(\theta_-, \theta^-, \xi_T)$, the sample path $\tilde{G}_{\omega,T}$ of \tilde{G}_T associated with ω is a smooth function of θ over Θ . Let G'_T and \tilde{G}'_T denote the derivative of G_T and \tilde{G}_T with respect to θ , respectively, then it is clear that $G'_T = O_{p.u.}(1)$ and $\tilde{G}'_T = O_{p.u.}(1)$.

Let

$$\begin{aligned} S_T^\bullet(\theta_-, \theta^-, \xi_T) &\triangleq \{ \omega \in S_T^\circ(\theta_-, \theta^-, \xi_T) \mid \sup_{\theta \in \Theta} |\tilde{G}'_T(\theta|\xi_T)| < d_T^{1/2}, \sup_{\theta \in \Theta} |G'_T(\theta)| < d_T^{1/2}, \\ &\text{and } \sup_{\theta \in \Theta} |C'_T(\theta)| < d_T^{1/2} \}. \end{aligned}$$

then for every $\omega \in S_T^\bullet(\theta_-, \theta^-, \xi_T)$, Equation (B.80) has the unique solution $\hat{\theta}_T$ when $T > T''$.

Thus, in this case, the FOC is a necessary and sufficient condition for minimum. By Lemma 3.2.8 and $\lim_{T \rightarrow \infty} d_T = +\infty$, we have $\lim_{T \rightarrow \infty} \Pr(S_T^\circ(\theta_-, \theta^-, \xi_T) \setminus S_T^\bullet(\theta_-, \theta^-, \xi_T)) = 0$. By (B.85), we have

$$\lim_{T \rightarrow \infty} \Pr(S_T^\bullet(\theta_-, \theta^-, \xi_T)) = 1. \quad (\text{B.86})$$

As the solution $\hat{\theta}_T$ of Equation (B.80) is local to θ_T , the root of $f_8(\theta, \delta_T, \phi_T)$ (as a function of θ), we consider equations

$$\tilde{G}_T(\theta|\xi_T) = d_T \left[\frac{\partial}{\partial \theta} f_8(\theta_T, \delta_T, \phi_T)(\theta - \theta_T) \right], \quad (\text{B.87})$$

and

$$-G_T(\theta) = d_T \left[\frac{\partial}{\partial \theta} f_8(\theta_T, \delta_T, \phi_T)(\theta - \theta_T) \right]. \quad (\text{B.88})$$

By Remark 3.2.15, (B.72) and (B.74), we know that $\frac{\partial}{\partial \theta} f_8(\theta_T, \delta_T, \phi_T) = \frac{1}{(1+\phi_T \theta_T)(1-\theta_T^2)} +$

$O(|\delta_T|)$. Thus, we further consider Equation (3.5).

It is clear that when T is sufficiently large, for every $\omega \in S_T^\bullet(\theta_-, \theta^-, \xi_T)$, there is a unique solution θ_T^+ , θ_T^\dagger , and θ_T^\diamond to Equation (B.87), (B.88), and (3.5), respectively. By elementary geometry, on $S_T^\bullet(\theta_-, \theta^-, \xi_T)$, we have⁴⁶

$$\hat{\theta}_T - \theta_T^+ = o_p(d_T^{-3/2}), \quad (\text{B.89})$$

$$\theta_T^+ - \theta_T^\dagger = O_p(d_T^{-1} \max\{|\delta_T|^{1/2}, d_T^{-1/2}\}), \quad (\text{B.90})$$

$$\theta_T^\dagger - \theta_T^\diamond = O_p(d_T^{-3/2}), \quad (\text{B.91})$$

and

$$\theta_T^\diamond - \theta_T = O_p(d_T^{-1}), \quad (\text{B.92})$$

which imply that for any $a \in \mathbb{R}$,

$$\lim_{T \rightarrow \infty} \Pr(d_T(\hat{\theta}_T - \theta_T) \leq a) = \lim_{T \rightarrow \infty} \Pr(d_T(\theta_T^\diamond - \theta_T) \leq a), \quad (\text{B.93})$$

when the limit distribution of either $d_T(\hat{\theta}_T - \theta_T)$ or $d_T(\theta_T^\diamond - \theta_T)$ exists. Moreover,

$$G_T(\theta_T^\diamond) - G_T(\theta_T) = O_p(d_T^{1/2} |\theta_T^\diamond - \theta_T|) = O_p(d_T^{-1/2}), \quad (\text{B.94})$$

which implies that for any $b \in \mathbb{R}$,

$$\lim_{T \rightarrow \infty} \Pr(G_T(\theta_T^\diamond) \leq b) = \lim_{T \rightarrow \infty} \Pr(G_T(\theta_T) \leq b), \quad (\text{B.95})$$

when the limit distribution of either $G_T(\theta_T^\diamond)$ or $G_T(\theta_T)$ exists.

⁴⁶By “on $S_T^\bullet(\theta_-, \theta^-, \xi_T)$ ”, we mean the restriction of r.v.’s on $S_T^\bullet(\theta_-, \theta^-, \xi_T)$, e.g., $(\hat{\theta}_T - \theta_T^+)1_{S_T^\bullet(\theta_-, \theta^-, \xi_T)}$. For notational simplicity, “ $1_{S_T^\bullet(\theta_-, \theta^-, \xi_T)}$ ” is suppressed in the expressions below. By (B.86), this restriction won’t affect the result.

It is obvious that on $S_T^\bullet(\theta_-, \theta^-, \xi_T)$,

$$\frac{d_T}{(1 + \phi_T \theta_T)(1 - \theta_T^2)^2}(\theta^\circ - \theta_T) = G_T(\theta_T^\circ). \quad (\text{B.96})$$

Then by (3.108) and (B.81), we have that

$$\begin{aligned} \lim_{T \rightarrow \infty} \Pr \left(\hat{\theta}_T - \theta_T \leq \frac{a}{d_T} \right) &= \lim_{T \rightarrow \infty} \Pr \left(\theta_T^\circ - \theta_T \leq \frac{a}{d_T} \right) \\ &= \lim_{T \rightarrow \infty} \Pr \left(G_T(\theta_T^\circ) \leq \frac{a}{(1 + \phi_T \theta_T)(1 - \theta_T^2)^2} \right) \\ &= \lim_{T \rightarrow \infty} \Pr \left(G_T(\theta_T) \leq \frac{a}{(1 + \phi_T \theta_T)(1 - \theta_T^2)^2} \right) \\ &= \Phi(a(1 - \theta_0^2)^{-3/2}), \end{aligned} \quad (\text{B.97})$$

where Φ is the standard normal distribution function. Thus, we have

$$d_T(\hat{\theta}_T - \theta_T) \xrightarrow{d} \mathcal{N}(0, (1 - \theta_0^2)^3), \text{ as } T \rightarrow \infty. \quad (\text{B.98})$$

More specifically, by (B.89), (B.90), (B.91), (B.94), and (B.96), we have

$$d_T \triangle \hat{\theta}_T = (1 + \phi_T \theta_T)(1 - \theta_T^2)^2 G_T(\theta_T) + O_p(\max\{|\delta_T|^{1/2}, d_T^{-1/2}\}). \quad (\text{B.99})$$

Thus, we have $\frac{1}{1 + \phi_T \theta_T} = \frac{1}{1 + \phi_T \theta_T} + O_p(d_T^{-1})$. By (3.75), we know that $f_1(\hat{\theta}_T, \delta_T, \phi_T) = 1 + O_p(|\delta_T|)$. By (3.81), we know that $f_2(\hat{\theta}_T, \delta_T, \phi_T) = (1 - \phi_T^2)\phi_T + O_p(\max\{d_T^{-1}, |\delta_T|\})$. Thus, by Theorem 3.2.11, we have

$$T^{1/2} \triangle \hat{\theta}_T = \phi_T(1 - \theta_T^2)^2 G_T(\theta_T) + (1 - \theta_T^2) C_T(\theta_T) + O_p(\max\{|\delta_T|^{1/2}, d_T^{-1/2}\}). \quad (\text{B.100})$$

Let $\underline{\theta}_T = (1 + \phi_T \theta_T)(1 - \theta_T^2)^2 G_T(\theta_T)$ and $\underline{\delta}_T = \phi_T(1 - \theta_T^2)^2 G_T(\theta_T) + (1 - \theta_T^2) C_T(\theta_T)$, by (3.110) and (3.111), we have

$$V_{\underline{\delta}_T, \underline{\theta}_T} \triangleq \text{Cov}((\underline{\delta}_T, \underline{\theta}_T)') = (1 - \theta_T^2) \begin{pmatrix} (1 + \phi_T^2) & \phi_T(1 + \phi_T \theta_T) \\ \phi_T(1 + \phi_T \theta_T) & (1 + \phi_T \theta_T)^2 \end{pmatrix}. \quad (\text{B.101})$$

By Lemma 3.2.3, and (B.99)–(B.101), we have (3.6), and the proof is complete. \square

Proof of Corollary 3.2.20. When $\delta_T = O(T^{-1/2})$, the conclusion holds by Theorem 3.2.11. When $\delta_T = O(T^{-1/2})$ is not true, by Theorem 3.1.3, we know that $\{\tilde{\xi}_T\}_{T=1}^\infty$ must be a $1/2\text{-}\langle\infty\rangle$ sequence or a convergent sequence with $\lim_{T \rightarrow \infty} \tilde{\xi}^* = (\delta^*, \theta^*)'$ and $\delta^* \neq 0$. The asymptotic distribution of $\hat{\xi}_T$ for the latter is given in Corollary B.1.5, which implies that the conclusion is true. Thus, we only need to check the case that $\{\tilde{\xi}_T\}_{T=1}^\infty$ is a $1/2\text{-}\langle\infty\rangle$ sequence.

In this case, by Case (3) of Theorem 3.1.2, $\Delta\hat{\theta}_T = O_p(d_T^{-1}) = O_p(T^{-1/2}\delta_T^{-1})$. Then $\Delta\hat{\theta}_T\delta_T = O_p(T^{-1/2})$. Thus, Theorem 3.2.11 implies that $\Delta\hat{\theta}_T = O_p(T^{-1/2})$, and the corollary is proved. \square

Proof of Corollary 3.2.19. This corollary immediately follows the Case (3) of Theorem 3.1.2. \square